Robust Texture Descriptors and Algorithms for Microcalcification Detection and Breast Density Estimation in Mammograms

A thesis submitted in partial fulfilment of the requirements
for the degree of
Doctor of Philosophy in Computer Science
at the
University of Canterbury
by

Haipeng Li

July 2021

This thesis is dedicated to my beloved parents,

Mr. Qingde Li and Mrs. Chunxian Wang.

ABSTRACT

This thesis proposes, implements and evaluates novel algorithms used for processing and analysing mammograms, with two distinct research aims: microcalcifications (MCs) detection and mammographic density (MD) classification. Mammography is the most widely used imaging technology for breast cancer screening, which outputs mammogram images for reading and interpretation by radiologists. The work presented in this thesis deals with robust and effective methods which can be used in computer aided detection systems to automatically detect diagnostic features and to assist radiologists in the interpretation of mammograms.

Accurate MCs detection is difficult due to their heterogeneous features: varied size, shape, and local contrast. This thesis presents a MCs detection framework with two stages: candidate MCs detection and false positive number reduction. In the first stage, a linear structure detector, multifractal features, and a support vector machine are applied to develop a MCs detector, which produces initial MCs detection results. In the second stage, a Weber's law-based multifractal measure is developed to enhance the texture features related to MC spots, and a patch-wise convolutional neural network (CNN) is proposed to classify the detected MC spots in the first stage into true positive or false positive groups. The false positive number is reduced significantly using the proposed scheme. A digital mammogram dataset, INbreast, is used to test the proposed detection framework, and experimental results show that the overall detection accuracy is higher than that obtained using other methods in the literature.

Mammographic density is a key biomarker of breast cancer, which indicates the risk of women developing breast cancer in the future, and MD evaluation plays an important role in preventing breast cancer. This thesis builds four robust and effective classification models for grouping mammogram images into different density categories, with an in-depth investigation of extracting and analysing relevant image features of mammograms. The image features used for MD evaluation in this thesis can be divided into three groups: cascaded image features, robust texture descriptors, and texture feature based spatial information.

For using the cascaded image features, this thesis presents a patch-wise classification method using multifractal spectrum, histogram information and Chi-square test statistic (model 1). In addition, histogram information based on both multifractal features and local binary patterns are incorporated to form a new feature set to be used in a density

classification model (model 2). An autoencoder network and principal components analysis are adopted to compress the feature space for improving the efficiency of the classifier.

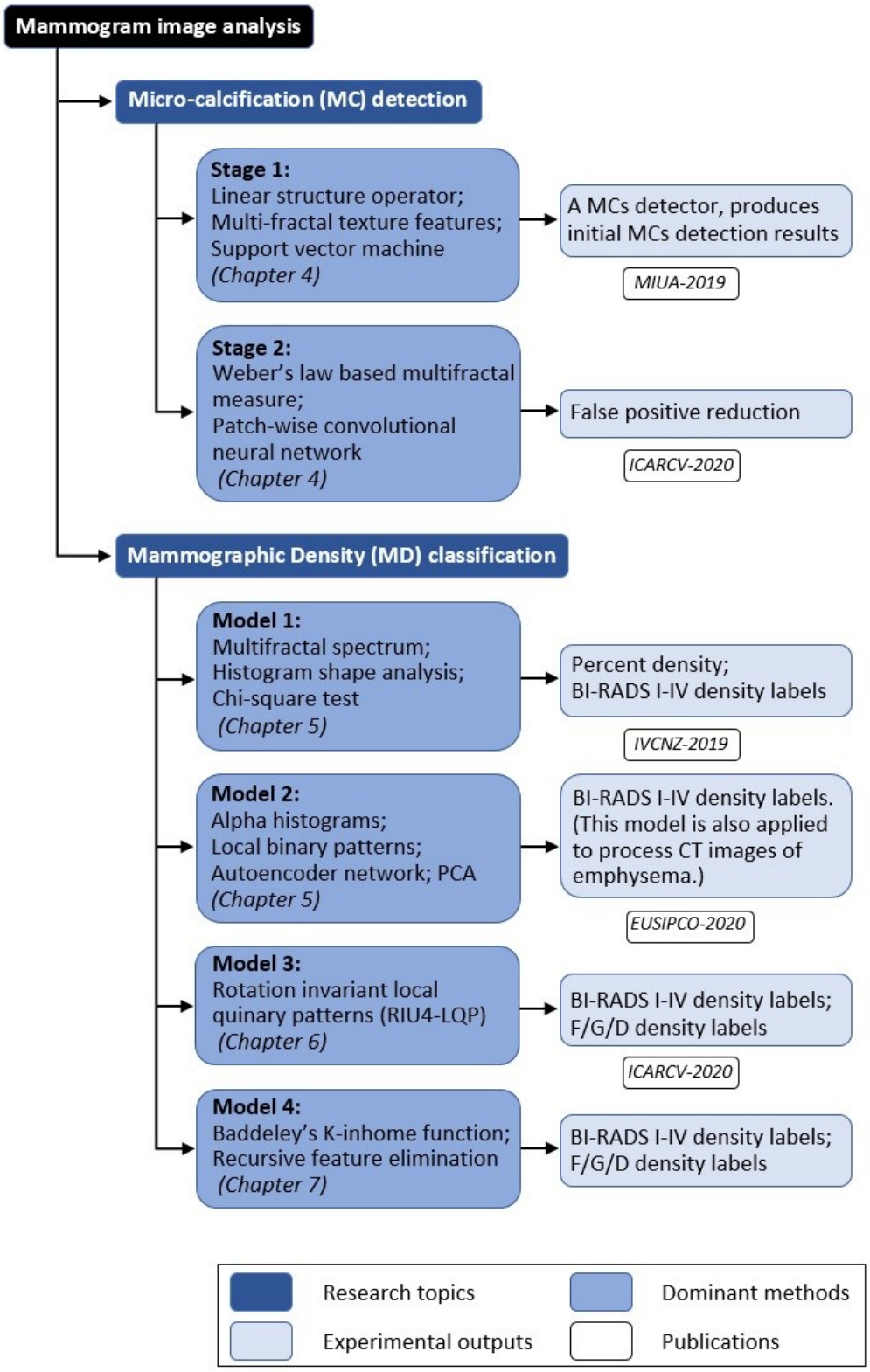
This thesis conducts a complete study based on binary encodings of local texture patterns, and proposes a MD classification method (model 3) by developing and testing novel texture descriptors. A robust rotation invariant texture descriptor is developed based on local quinary patterns using high transition numbers.

This thesis also explores spatial distribution characteristics of feature points based on the above novel texture descriptor to capture rich image representations and to improve the classification accuracy. Baddeley's K-inhom method is employed to describe the spatial information of feature points in mammograms and a new texture feature vector, K-spectrum, is constructed and used in the classification model (model 4). The proposed classification method is tested on two mammogram datasets, INbreast and MIAS. Quantitative assessment and comparative analysis are given, which demonstrate the superiority and robustness of the proposed model.

Finally, this thesis summarises the main contributions and research accomplishments, and also outlines extensions and future directions.

Keywords*—*Mammogram, Breast cancer, Microcalcification detection, Breast density classification, Multifractal analysis, Texture feature extraction, Image enhancing technique, Convolutional neural network, Autoencoder network, Local binary patterns, Local quinary patterns, Rotation invariant uniform encoding, Spatial distribution analysis, Feature selection

GRAPHICAL ABSTRACT



ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my supervisor, Professor Ramakrishnan Mukundan, who gave me the precious opportunity of pursuing my PhD study, and the trust and support throughout this research journey. His patience, vision, sincerity and immense knowledge have deeply inspired me. Without his guidance and persistent help this dissertation would not have been possible. It was a great privilege and honour to work and study under his guidance. I would also like to thank my associate supervisor Dr. Shelley Boyd, Canterbury breastcare, St. George's Hospital, for her encouragement, timely suggestions, valuable feedback and professional medical opinions which contributed a lot to this thesis.

I am grateful to the Department of Computer Science and Software Engineering for providing a friendly and comfortable research environment, and to all the technical and administrative staff for their help and support.

My appreciation also extends to my laboratory friends and colleagues, Alan, Ori, Tieta, Prerna, Dibash, Sarmad, and Anthony, for their constant encouragement and I will never forget all the fun we have had in the last three years. I particularly express my thanks to my friends in China, Feng Li, Qian Li, and my lovely niece Shiqi Li, who share my happiness and sadness in those video calls. Also I thank my friends Julio and Ashu, and will not forget those travelling experience with you together.

Finally, I am extremely grateful to my parents for their love, caring, understanding and sacrifices for educating and preparing me for my future.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

ACC	Accuracy
ACR	American College of Radiology
ANLS	Automatic Non-Linear Stretching
AUC	Area under the ROC Curve
BIF	Basic Image Features
BI-RADS	Breast Imaging-Reporting and Data System
CA	Classification Accuracy
CAD	Computer Aided Detection
CC	Craniocaudal
CNN	Convolutional Neural Network
CT	Computerized Tomography
DAG-SVM	Directed Acyclic Graph Support Vector Machine
DDSM	Digital Database for Screening Mammography
DICOM	Digital Imaging and Communications in Medicine
DPS	Dominant Patterns Selection
FC	Fully Connected
FCN	Fully Convolutional Network
FFDM	Full Field Digital Mammogram
FIR	Feature Importance Ranking
FP	False Positive
FROC	Free-Response Receiver Operating Characteristic Curve
GLCM	Gray-Level Co-Occurrence Matrix
GT	Ground Truth
LBP	Local Binary Patterns
LGA	Local Greylevel Appearance
LQP	Local Quinary Patterns

LTP	Local Ternary Patterns
MC	Micro-calcification
MD	Mammographic Density
MIAS	Mammographic Image Analysis Society
MLO	Mediolateral Oblique
OAA	One Against All
PCA	Principal Component Analysis
PD	Percent Density
PGM	Portable Gray Map
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
RFE	Recursive Feature Elimination
RIU	Rotation Invariant Uniform
ROC	Receiver Operating Characteristic Curve
ROI	Region of Interest
SCC	Six-Class-Categories
SD	Standard Deviation
SFM	Scan Field Mammogram
SGD	Stochastic Gradient Descent
SK	Skewness
SPCNN	Simplified Pulse Coupled Neural Network
SVM	Support Vector Machine
TNR	True Negative Rate
TP	True Positive
TPF	True Positive Fraction
TPR	True Positive Rate
VAS	Visual Analogue Scale

LIST OF SYMBOLS FOR FUNCTIONS

$c(x, r)$	Correction function of edge effects
$f(\alpha)$	Fractal dimension
$g(x)$	The intensity value of a pixel
$n(\alpha)$	The number of points with the Hölder exponent value of α
$K_{inhom}(r)$	Baddeley's K-inhom function
$K_{pois}(r)$	Expected reference value
$L(x)$	The average grey-level on a line
$LBP(c, R, P)$	Local binary pattern encoding function
$LQP(c, R, P)$	Local quinary pattern encoding function
$N(x)$	The average intensity in a local background
$S(x)$	Line strength signal
$T(\cdot)$	The number of spatial transitions
$\lambda(x)$	Intensity function
$\hat{\lambda}(x)$	The estimation of $\lambda(x)$
$\mu_w(p)$	Multifractal measure function
d	Deviation between the observed $K_{inhom}(r)$ value and the reference value $K_{pois}(r)$
$\bar{d}$	The mean of a number of d values
p	A central pixel in a square window
r	Distance measure
x	One pixel in an image
P	The number of neighbour points used in LBP
R	Radius of a local region
W	Study area
PD	Breast percentage density

SD	Standard deviation
SK	Skewness
α	Singularity coefficient (Hölder exponent)
μ	The mean of a numeric vector
τ	The threshold value used in LQP function
w	The edge length of a square window
χ^2	Chi-square statistic

1 INTRODUCTION

Mammography is the most widely used imaging technology for breast cancer screening, which produces mammogram images for reading and interpretation by radiologists. Computer aided detection systems that use image analysis and machine learning algorithms can significantly improve radiologists' workflow and produce higher detection accuracy. This thesis addresses novel and robust computerised algorithms to automatically detect breast anomalies and analyse image characteristics related to breast cancer in mammograms, with the aim of offering a secondary perspective to assist radiologists' work. Concretely, this thesis focuses on two different tasks: microcalcifications (MCs) detection and mammographic density (MD) classification. An effective MCs detection framework is proposed for detecting individual MC spots and locating their positions in mammograms, which contains two stages: candidate MCs detection and false positive reduction. This thesis develops texture feature enhancement methods to extract image information required for recognizing MC spots, and a convolutional neural network based classification model to reduce the false positive number. For evaluating breast density, this thesis proposes four different classification models with the aim of classifying mammograms into target density groups, and different image feature extraction methods are explored and investigated in these models, including cascaded image features, texture descriptors, and spatial information based on novel texture features. This chapter briefly discusses the research motivation, research aims, the methodology used in this work, and the main contributions of this thesis.

1.1 Breast Cancer and Mammograms

Breast cancer is one of the most common cancers in women with over 1.5 million females worldwide diagnosed each year. Nearly 627,000 patients died of breast cancer in 2018, representing approximately 15% of all cancer deaths among women (WHO, 2019). In New Zealand, breast cancer is the most common cancer for women, with more than 3,000 diagnosed cases and more than 600 deaths every year (Breast Cancer Foundation New Zealand, 2019). Due to the absence of an effective cure, early detection through routine screening plays an important role in preventing breast cancer deaths and reducing the associated morbidity. Currently, mammography is the widely accepted imaging method used for routine breast cancer screening in many countries. It can detect about 80%─90% of breast cancer cases before development of symptoms (American Cancer Society, 2017).

Microcalcifications are small deposits of calcium in the breast and clustered MCs can be a primary indicator for both benign and malignant pathology. However, accurate detection of MCs in mammograms is challenging and difficult. The breast contains various tissue components, such as glandular regions, vessels, and connective tissue. If there is a large amount of glandular tissue, the mammogram image can be very bright or white, making small MCs poorly visible. In addition, MCs can have heterogeneous properties, including varied size (ranging from 0.1mm to 0.5mm), shape (e.g. ring, round, granular, linear), and poor local contrast.

Mammographic density is a critical bio-marker which indicates the possibility of developing breast cancer in the future for women. High breast density is caused by a high percentage of fibro-glandular tissue and reduces the effectiveness of mammography screening. Related research work shows that women with extremely dense breasts could suffer four to six-fold higher risk of developing breast cancer than other females with low breast density (Boyd et al., 1982). Breast density measurements have been proposed as a screening tool to help identify and target women who could benefit from tailored screening options such as increased (or decreased) screening interval or supplemental screening via alternative modalities (McLean & Stone, 2018).

This thesis aims to develop automated methods for analysing mammogram images with two different tasks: microcalcifications detection and mammographic density classification.

1.2 Research Motivation

Although some research work has been done in the literature for MCs detection, most work focuses on producing a qualitative label for MC clusters, such as the existence of a cluster in a designated region, without giving the location information of individual MCs. The first task (i.e. MCs detection) in this thesis is motivated by the following factors:

- MCs can be a primary indicator for both benign and malignant pathology.
- Accurate MC detection is a challenging work. Visual assessment in mammograms by radiologists may lead to a 10%─30% rate of undetected lesions.
- Current methods do not pay enough attention to individual MC spots detection, which offer important information to radiologists.
- Successful applications of texture enhancing methods and convolutional neural networks in the field of medical image processing make it possible to develop effective MCs detection algorithms.

Breast density is a critical bio-marker which indicates the risk of developing breast cancer in the future for women. Dense tissue areas in mammograms also cause the 'masking' effect leading to reduced sensitivity when radiologists visually assess related breast lesions or early signs of cancer, such as lumps and calcification clusters. The second task (i.e. MD classification) in this thesis is motivated by the following points:

- Breast density estimation depends on radiologists' visual assessment, which is time-consuming.
- Density assessment is not always consistent, with only 57% inter-reader agreement and 77% intra-reader agreement (Volpara health, 2020).
- There is a need for improving the density classification accuracy, with only few reported results in the literature surpassing 80% accuracy.
- In recent studies, the use of extracted texture features in mammograms for breast density classification has obtained promising results. However, these methods may lead to problems such as high feature dimensionality and the absence of spatial information of feature points.

1.3 Research Questions and Aims

The main research questions addressed in this thesis are listed in this section.

For **MCs detection**, this thesis considers the following questions:

- What are the main features of MC spots in mammograms? Are there any uniform morphological features that could be used to develop MCs detection algorithms?
- What is the current state of research on MCs detection?
- What are the main challenges in recognizing MC spots in mammograms?
- How critical are false positive and false negative values in MC detection using existing methods?
- If the false positive number or false negative number is high, what scheme can be used to reduce those numbers?

By considering the above research questions, this thesis develops a novel MCs detection framework with two stages: candidate MC spots detection and false positive reduction. In the first stage, this study aims to develop a MCs detector by enhancing MCs related texture features to recognize MCs spots. The proposed MCs detector produces a high true positive rate but also sees a high false positive number (the last research question above). Therefore, the research aim in the second stage is to reduce the false positive number and improve the overall detection accuracy.

For **MD classification**, the key research questions addressed by the thesis are as given below:

- What image characteristics of dense tissue areas in the breast region should be used for classifying mammograms into different density categories?
- What kind of image features have been used in the classification task? What are the disadvantages of those methods?
- Is it possible to build a deep learning model to classify breast density?
- Which classifier is most appropriate for this classification work?
- How to select the optimal image information from the initial feature set for improving the classification performance?
- Is histogram information sufficient to produce desirable classification results when using image texture features?
- How to extract spatial information based on texture features? Can spatial information contribute to the improvement of the classification accuracy?

- What factors affect the measurement of classification results? How to conduct a comprehensive and fair comparison between different methods?

This thesis proposes four different classification models with the aim of extracting various image features which can be used to improve the classification performance. Models 1 and 2 aim to cascade multiple image features to evaluate mammographic density. Model 3 focuses on developing a novel texture descriptor to extract effective texture features and also maintain a manageable number of feature number. Model 4 uses spatial information based on texture features for improving the classification accuracy. Meanwhile, different feature selection schemes are used and compared in the proposed classification model.

1.4 Research Methodology

Based on the research questions and aims discussed in the previous section, this thesis proposes corresponding methods to achieve the research goals. Since the MCs detection task is divided into two stages and breast density classification work is done through four different models, the methodology developed in this thesis is introduced separately in these part and a diagrammatic overview of the following stages/methods is given in the graphical abstract at the beginning of the thesis.

MCs detection stage 1: This study proposes a MCs detector to obtain the initial detection results. By considering the heterogeneous features of MCs, multifractal analysis with four different local measures is used to enhance the texture features of regions where microcalcifications are present. In addition, a linear structure detector is adopted in this stage to distinguish glandular tissue (in linear shape) from the MCs. A support vector machine is used to analyse multifractal features and intensity information of candidate MC spots to produce the initial detection results.

MCs detection stage 2: Since the initial detection results produce high false positive values, this stage uses the detection results as the input and develops methods to reduce the high false positive number. A Weber's law based multifractal measure is proposed to extract novel texture features for further enhancing texture patterns of MC spots. A CNN based classification model is built in this stage to classify the MC candidates to true positive or false positive groups, reducing the false positive number. In addition to inputting original mammogram patches to the CNN classifier, texture enhanced patches are also input to the classification model for improving the classification accuracy.

MD classification models 1 and 2: In the first two models, this study considers cascading multiple image features to distinguish between dense and fatty tissue areas for classifying mammograms into different density categories. Model 1 explores multifractal spectrum which integrates a number of fractal dimensions along with texture feature values to describe the global image features based on original patches. Model 2 uses different multifractal measures to obtain various texture features, and concatenates histograms based on multifractal features and local binary patterns to form an initial feature set. An autoencoder network and principal component analysis are used to optimise the cascaded features.

MD classification model 3: This model particularly focuses on developing a novel texture descriptor for extracting robust and effective texture features used in this classification task. After comparing and analysing the currently used texture feature extraction methods, local quinary patterns (LQP) are used as the prototype and improvement strategies are proposed to extend LQP. This study applies a rotation invariant method based on LQP and also considers a wider range of transition numbers to develop the rotation invariant uniform LQP (RIU4-LQP) with superior feature representation capabilities.

MD classification model 4: Based on the proposed texture descriptor in model 3, this model further explores spatial information of texture features. Baddeley's K-inhom method is used to describe the spatial distribution characteristics of feature points, based on which a new feature vector called K-spectrum is developed and used in the classification task. In addition, this model uses and compares three different feature selection methods to optimise the initial feature space, for filtering the optimal feature set and improving the classification accuracy.

1.5 Thesis Contributions

The main contributions of this thesis are outlined below.

In the area of MCs detection algorithms, this thesis proposes a multifractal based framework for enhancing texture features and reducing false positives. A CNN based classifier is developed and evaluated using the INbreast mammogram dataset and desirable MCs detection results are achieved. Related work is published in the following papers:

1) Li H., Mukundan R., & Boyd S. (2019). A Novel Application of Multifractal Features for Detection of Microcalcifications in Digital Mammograms. *Paper presented at the Medical Image Understanding and Analysis (MIUA-2019)*, University of Livepool, UK. https://doi.org/10.1007/978-3-030-39343-4_3
2) Li H., Mukundan R., & Boyd S. (2020). An Improved Micro-Calcification Detection Algorithm Using a Novel Multifractal Texture Descriptor and CNN. *The 16th International Conference on Control, Automation, Robotics and Vision (ICARCV-2020)*, Shenzhen, China. https://doi.org/10.1109/ICARCV50220.2020.9305306

In the area of MD classification algorithms, this thesis develops four classification models and investigates different image feature extraction methods as described in the previous section. One of the models has also been successfully applied to process emphysema CT images for classifying emphysema subtypes. This research work has produced the following papers:

3) Li H., Mukundan R., & Boyd S. (2019). Breast Density Classification Using Multifractal Spectrum with Histogram Analysis. *Paper presented at the 2019 International Conference on Image and Vision Computing New Zealand (IVCNZ-2019)*, University of Otago, New Zealand. https://doi.org/10.1109/IVCNZ48456.2019.8961037
4) Li H. and Mukundan R. (2020). Robust Texture Features for Emphysema Classification in CT Images. *The 28th European Signal Processing Conference (EUSIPCO-2020)*, Amsterdam, Netherlands. https://doi.org/10.23919/Eusipco47968.2020.9287761
5) Li H., Mukundan R., & Boyd S. (2020). Robust Texture Features for Breast Density Classification in Mammograms. *The 16th International Conference on Control, Automation, Robotics and Vision (ICARCV-2020)*, Shenzhen, China. https://doi.org/10.1109/ICARCV50220.2020.9305431

1.6 Thesis Organisation

The remaining parts of the thesis are organized as follows:

Chapter 2 presents the literature review in which relevant methods used in the two research tasks are studied and summarised. The approaches related to MCs detection in mammograms are first reviewed with the aim of investigating MCs features which are commonly considered and used in other work. Then the literature review on breast

density classification is given. By comparing the reported methods, their advantages and drawbacks are summarised, based on which research questions and aims are proposed in this thesis.

Chapter 3 introduces the materials and main methods used in this thesis. Two mammogram datasets are used to test the proposed methods in the thesis, and their basic information is given in this chapter. The pre-processing operations used in the proposed two tasks, and relevant methods are introduced. In addition, the main methods developed in the two tasks are described.

Chapter 4 elaborates on the proposed MCs detection framework. The two stages of this framework are introduced by presenting the developed methods, experimental settings, and results evaluation.

Chapter 5 gives detailed information of the development of the breast density classification models 1 and 2. Multiple image features used in the two models including multifractal spectrum, texture features based histograms, and local binary patterns are introduced in this chapter. The design of the two classification models using the Chi-square test statistic and an autoencoder network are detailed as well.

Chapter 6 presents the classification model 3 with a novel texture descriptor RIU4-LQP. Two main considerations, namely, rotation invariance and a wider range of transition numbers, are explained and the effectiveness of the extended descriptor is demonstrated in experiments. Two mammogram datasets are used to test this classification model.

Chapter 7 describes the MD classification model 4. Based on the developed texture descriptor in model 3, this model focusses on characterising spatial information related to RIU4-LQP features. This chapter develops methods for extracting spatial characteristics of features, and introduces three feature selection schemes. Quantitative evaluation and comparative analysis are conducted based on experimental results, showing the differences in classification performance by using different image features.

Chapter 8 summarises the work reported in this thesis with the main contributions and future work.

2 BACKGROUND RESEARCH

In this chapter, research work reported in the literature on microcalcification detection and mammographic density evaluation in mammograms are reviewed. By investigating the reported methods and experimental results, this chapter summarises their merits and disadvantages, based on which some important research questions in this thesis are proposed and improvement schemes are developed in this study. In addition, through literature review, this chapter collects the metrics that are commonly used for evaluating experimental results, which provides the basis for comparative analysis under the same measurements between the proposed methods in this study and other approaches in the literature. Since this thesis develops one MCs detection framework and four MD classification models which relate to three types of image feature extraction methods, the relevant literature review is arranged in different sections. This chapter first discusses applications of automated methods in mammogram analysis and then presents the literature review in subsequent sections to correspond to the two research works and different method types. A brief summary is attached at the end of each review section to give a clear distinction and relation between the proposed methods in this thesis and the reviewed approaches in each section.

2.1 Automated Methods in Mammography Screening

(Kerlikowske et al., 2000)(Sampat, Bovik, & Whitman, 2008) concluded that reading and interpreting suspicious regions in mammograms by radiologists is a repetitive and fatiguing task, leading to a 10%–30% rate of undetected lesions. To decrease this rate, computer-aided detection (CAD) systems have been developed in the

past two decades to assist radiologists in the interpretation of medical images (Alasadi & Al-Saedi, 2017) (Saad, Khadour, & Kanafani, 2016) (Singh & Kaur, 2018). Although using automated methods to detect breast lesions (e.g. MCs) may improve the sensitivity of mammographic screening, this may be offset by a higher recall rate and the potential of over diagnosis. Research work in (Ribli, Horvath, Unger, Pollner, & Csabai, 2018) (Lehman et al., 2015) (Fenton et al., 2011) reported that the benefits of using CAD are controversial. Some clinical trials in different countries showed that CAD technologies do not improve the performance of radiologists' work.

(Volpara health, 2020) reported that mammographic density evaluation by radiologists' visual assessment is not always consistent, with only 57% inter-reader agreement and 77% intra-reader agreement. In recent years, commercial software tools have been made available to automatically evaluate mammographic density by producing quantitative or qualitative assessment results. The effectiveness of automatically evaluating breast density in clinical applications has not been established. A recent case-control study in (Astley et al., 2018) indicated that compared to using automated methods such as Cumulus (Byng, Boyd, Fishell, Jong, & Yaffe, 1994), Volpara (Highnam, Brady, Yaffe, Karssemeijer, & Harvey, 2010), and Quantra (Ciatto et al., 2012), the visual assessment by Visual Analogue Scale (VAS) is the strongest predictor of screen-detected cancer after adjustment for classical risk factors. In (Brandt et al., 2016), two commercial software tools were used to compare their performance with visual evaluation by radiologists using the BI-RADS criterion (Sickles et al., 2013), and a disagreement rate up to 14% was found when classifying mammograms with dense tissue.

The above information indicates that current detection and classification methods used for analysing mammograms have not well satisfied the requirements of clinical applications or improved radiologists' workflow significantly. Therefore, further work is needed for improving and optimising the existing methods.

2.2 Microcalcifications Detection

MCs are one of the important image-based biomarkers for diagnosing breast lesions (Naseem et al., 2015), with approximately 50% of diagnosed cases at this early stage presenting MCs (Scimeca et al., 2014). This section specifically reviews MCs detection methods proposed in recent years with the aim of offering the context related to the

research work in this thesis and also identifying any gaps existing in the currently used methods.

(Guo et al., 2016) adopted a simple nonlinear function to modify the contourlet coefficients for enhancing MCs in mammograms and removing noise. After retaining the significant information, a suspicious calcification area was obtained first. A non-linking simplified pulse coupled neural network (SPCNN) method was used in their work for detecting MC clusters, and a density maximum principle was used to remove isolated points and improve the true positive rate. MIAS (J Suckling et al., 2015) and JSMIT datasets, and 20 mammograms from a local hospital were used to validate the proposed method. Detection results were reported with specificity of 94.7%, sensitivity of 96.3%, and AUC of 0.97.

(Mehdi et al., 2017) combined a spatial automatic non-linear stretching (ANLS) scheme and Shannon entropy based wavelet coefficient thresholding (SE_WCT), by which MCs details were enhanced and used to conduct the MCs detection work. The Mini MIAS database was used to test their method and experimental results were evaluated via FROC curves. The method obtained AUC of 0.92, sensitivity of 97% and 0.48 false positives per image.

(Ciecholewski, 2017) presented a method for detecting and segmenting MCs in mammograms by using morphological transformations and watershed segmentation methods. With the proposed method, MC regions in ROIs were detected morphologically and an approximate area was determined, followed by improving contrast and reducing noise. 200 ROIs cropped from mammograms in the DDSM dataset were used in their experiments. The main limitation of this work was that MCs-contained ROIs were manually segmented by radiologists but not automatically detected.

(Liu, Mei, Liu & Hu, 2015) integrated the possibilistic fuzzy c-means (PFCM) clustering algorithm and weighted support vector machine (WSVM) for detecting MCs in mammograms. 22 out of 51 extracted features were selected and used in the training and testing process, producing the optimal experimental results with sensitivity of 92% and 2.3 false-positive clusters per image. The proposed method obtained an AUC value of 0.8676.

(Stojic, Rejin, & Rejin, 2006) (Stojic & Reljin, 2010) used multifractal analysis to perform MCs segmentation in mammograms. To obtain a better segmentation, image pixels from a specific range of fractal values were selected and presented as white points on a black background. Successive morphological closing and opening procedures were applied as post-processing steps to generate the final MCs segmentation results. There was no quantitative analysis of experimental results and the ROIs containing MCs were selected manually from the MIAS dataset using the ground truth information.

(Alasadi & Al-Saedi, 2017) proposed two dimensional discrete wavelet transform to detect MCs. Different wavelet families, such as Daubechies, Symlets, and Coiflet were considered and compared for the same detection task, and Daubechies showed a better performance. However, only 29 mammograms in the MIAS dataset were used to test the proposed method.

(Cea, Yang, & Nishikawa, 2018) presented a new feature extraction approach to reduce the influence of false positives in the detected MCs in a cluster. Their method focused on two broad categories of features: cluster features for describing MCs' spatial distribution patterns and image features related to individual MCs. A private mammogram dataset containing 186 images was used in their experiment and an improved classification performance was obtained.

(de Cea, Nishikawa, & Yang, 2018) proposed outlier detection algorithms to identify MCs which were treated as statistical outliers in the detection set. They investigated two outlier detection methods, Mahalanobis distance outlier detector and Stochastic Neighbour Graph (SNG) outlier detection method, to develop an adaptive decision scheme for MCs detection. The evaluation results demonstrated that the proposed method improved the detection accuracy over the traditional approach of uniform thresholding.

(Wang, Yang, & Nishikawa, 2013) concluded that linear structures are a major source of false positives when detecting clustered MCs in mammograms. They developed a linear structure detection procedure together with a dual-thresholding scheme to separate the linear structures from tissue background in mammograms. Their experimental results demonstrated that the proposed method reduced false positives effectively after testing 200 mammograms in a private dataset.

(Domingues & Cardoso, 2014) used a Bayesian Surprise method to detect MCs in digital mammograms. By computing the Bayesian Surprise values related to a patch and its neighbourhood, a region could be detected as a MC spot if its value exceeds a threshold. A public mammogram dataset, INbreast, with pixel level annotations for individual MC spots by radiologists, was used to test the proposed method. Experimental results were analysed by the free-response receiver operating characteristic (FROC) curve, with detection sensitivity of 60.3% when the average false positive number is 108 per image.

(Wang & Yang, 2018) developed a context-sensitive deep neural network to detect individual MCs. This method considered local image features related to MCs and their neighbourhood's tissue background. There were 521 SFM images and 188 FFDM images collected for testing the proposed method, and 27,022 MCs manually identified by experienced radiologists used as ground truth. For evaluating the experimental results, FROC analysis was conducted, with TPF at 80% when the average FP rate is 1.03 FPs/cm^2.

Section summary: The review presented above shows that MC detection algorithms commonly try to enhance image contrast before extracting relevant image features, based on which MCs are identified from tissue background. Some research work did not detect individual MCs, instead detected MC clusters due to the lack of MCs ground truth. However, accurate MC cluster detection is based on individual MCs identification, which can offer more concrete information to radiologists. In addition, false positives are a common occurrence in the outputs of MCs detectors, leading to a reduction in the overall detection accuracy. Therefore, this thesis proposes a complete MCs detection framework with two stages in **Chapter 4**, by developing novel image feature enhancing techniques to detect MCs and designing a CNN based classifier to reduce FPs.

2.3 Heterogeneous Features for Density Classification

Mammographic density classification aims to measure the amount of radiodense tissue (i.e. fibro-glandular tissue) in mammograms for indicating the risk of developing breast cancer in the future. This section reviews the literature focusing on extracting heterogeneous features of fibro-glandular tissue to classify breast density.

(Mario, Mislav, & Krešimir, 2012) extracted multiple features based on image intensity, histograms, and GLCM to classify MD. Wrappers were used for feature selection and improving classification results.

(Oliver et al., 2015) combined intensity, texture, and morphological features to classify pixels in mammograms into two categories (fatty and dense) using an SVM classifier.

(Subashini, Ramalingam, & Palanivel, 2010) used statistical features including mean, standard deviation, smoothness, third moment, uniformity, and entropy to classify mammograms into three density categories.

(Qu et al., 2020) proposed a fuzzy-rough refined image processing method to enhance local image regions and to extract GLCM based statistical features for classifying mammographic density.

(Li et al., 2018) extracted 137 pixel-level features containing intensity, GLCM, and morphological features, to group pixels into fatty or dense classes. A regression analysis method (LASSO) is used in their work for performing variable selection and regularization.

(Tzikopoulos, Mavroforakis, Georgiou, Dimitropoulos, & Theodoridis, 2011) extracted 21 features based on intensity and fractal texture features, and SVM was used to classify MIAS mammograms into 3 categories.

(Muhimmah & Zwiggelaar, 2006) used multi-resolution histograms to analyse texture features, and mammograms were classified into 3 density categories by a Directed Acyclic Graph (DAG)-SVM classifier.

(Chen, Denton, & Zwiggelaar, 2011) investigated 5 texture feature sets separately, including LBP, Local Greylevel Appearance (LGA), Textons (MR8 texton, and image-patch texton), and Basic Image Features (BIF). Their experimental results showed that image-patch texton features produced a higher classifying accuracy for 4 BI-RADS categories classification.

(George, Rampun, Denton, & Zwiggelaar, 2016) proposed a multi-scale blob detection method to recognize the fatty and dense tissue present in mammograms. This method was used to analyse MIAS mammograms and experimental results revealed some initial relations between the BI-RADS density category and the average relative tissue (fatty and dense) area in mammograms.

(Zheng et al., 2015) used a lattice-based approach to extract statistical and structural features for analysing parenchymal texture in mammograms.

Section summary: The methods reviewed in this section particularly focus on cascading multiple image features to collect as much information as possible to improve the classification accuracy. Based on this methodology, this thesis develops two different classification models in **Chapter 5 (model 1 and 2)** by exploring multifractal spectrum, texture features based statistical information, and local binary patterns to extract relevant image features.

2.4 Deep Learning Based Features

Deep learning based methods have been used recently to analyse medical images for both classification and segmentation tasks, with promising results. This section gives a review of the literature using deep learning based methods to automatically extract image features for the MD classification work.

(Mohamed et al., 2018) proposed a CNN model based on AlexNet to distinguish between two BI-RADS categories ('scattered density' and 'heterogeneously dense') which often leads to disagreements in radiology assessments (Berg, Campassi, Langenberg, & Sexton, 2000).

(Ahn, Heo, Jin, & Kim, 2017) designed a CNN architecture to learn relevant features from a multitude of sub-images and to classify them into dense and fatty tissue.

(Lee & Nishikawa, 2018) used a fully convolutional network (FCN) to segment the breast region and fibro-glandular areas with the aim of estimating percentage density.

(Li et al., 2018) trained a deep convolutional neural network to classify mammographic pixels into fatty class or dense class. A probability map of breast density was generated and used to estimate percentage density.

(Kallenberg, 2016) proposed an unsupervised deep learning model to segment dense breast regions in mammograms.

Section summary: By reviewing the recently proposed methods for evaluating MD, we can find that CNN based methods were usually proposed to deal with binary classification tasks (e.g. fatty or dense tissue) rather than multi-class grouping (e.g. BI-RADS categories). One of the main limitations of applying deep learning methods to classify MD is that it requires a huge number of training images with accurate

annotations by clinicians (Hamidinekoo, Denton, Rampun, Honnor, & Zwiggelaar, 2018). This thesis aims to classify mammograms into multiple density categories (3 or 4), and therefore does not directly consider using CNN to classify mammograms but applies an autoencoder network to optimise the extracted features in the proposed classification **model 2 in Chapter 5**.

2.5 Texture Descriptors

The image quality of mammograms is susceptible to the radiation dose used in the screening routine. Therefore, image processing methods based on original intensity information do not produce desirable results; instead, texture analysis methods that extract local image structure information are more effective in capturing image features. This section investigates the use of texture descriptors for analysing medical images and reviews the literature on texture feature extraction methods for MD classification.

(Ojala, Pietikäinen, & Harwood, 1996) proposed local binary patterns (LBP) to describe image texture patterns. Due to its simplicity and efficiency, LBP has been studied widely and a few variants were also proposed for extracting texture features and classifying medical images.

(George & Zwiggelaar, 2019) extended LBP to elliptical LBP (ELBP) and mean-elliptical LBP (M-ELBP) by considering various neighbourhood topologies and different local region scales. M-ELBP presented a better classification result (77.38 ± 1.06) on the MIAS dataset.

(Tan & Triggs, 2010) proposed local ternary patterns (LTP) using a 3-value encoding approach compared to a 2-value encoding of LBP. (Peng et al., 2017) extended LTP to rotation invariant uniform LTP (RIULTP) and Weber-based RIULTP for classifying pulmonary emphysema in CT images into different subtypes.

(Rampun, Morrow, Scotney, & Winder, 2017) extracted LTP based texture features to classify MIAS mammograms into 4 BI-RADS categories.

(Nanni, Lumini, & Brahnam, 2010) proposed local quinary patterns (LQP) by extending LBP from a binary encoding system to a 5-value encoding system, and used three different medical image classification tasks to test this new texture descriptor.

(Rampun, Scotney, Morrow, Wang, & Winder, 2018) extended LQP with multi-resolution and multi-orientation schemes to classify MD. Their experimental results

showed that the use of LQP based texture features improved MD classification accuracy.

(Rampun, Morrow, Scotney, & Wang, 2020) recently introduced a local septenary patterns (LSP) method by using a 7-value encoding approach to further improve the MD classification performance. Their experimental results demonstrated that classification accuracy was slightly improved by LSP compared to LQP (80.5 ± 9.2 vs 80.1 ± 10.5 on INBreast), but the statistical test indicated that the results difference is not statistically significant (p = 0.45).

Section summary: Through the literature review related to the use of texture descriptors for classifying MD, we can see that extracting effective texture features in mammograms has been an active research field in recent years, with related research work developing various texture descriptors and reporting competitive classification results. However, the currently used texture descriptors in the literature still have the problems of high feature dimensionality, lack of effective feature selection methods, and absence of any spatial information, which restrict further improvement of the classification performance. These problems are analysed and discussed in this thesis, with a novel texture descriptor developed and used in the proposed classification **model 3 (Chapter 6)** and a new spatial information based feature vector designed in **model 4 (Chapter 7)**.

2.6 Chapter Summary

This chapter reviews literature relevant to two research topics: MCs detection and MD evaluation. Based on the literature review in section 2.2, we list representative MCs detection methods in Table 2.1, which gives a direct comparison on dataset, research aims, results evaluation metrics, and reported detection results. We observe the following important points from Table 2.1:

1) Public and private datasets containing different ground truth information are used in related research work, with varying detection performance reported, which make it difficult to conduct comparative analysis between different methods.
2) Research work focusing on individual MCs detection has not been done commonly due to the absence of annotations by radiologists in publicly available datasets.

3) Free-Response Receiver Operating Characteristic (FROC) analysis is the commonly used results evaluation method when detecting MCs, which reflects the detection performance comprehensively.

In **Chapter 4**, a MCs detection framework is developed and tested using the INbreast dataset (introduced in **Chapter 3**), which is a public mammogram dataset and contains pixel level annotations for individual MC spots. The detection results in this thesis are evaluated by FROC analysis and compared with other methods.

In this chapter, different image feature extraction methods used for classifying breast density have been grouped into 3 parts and have been reviewed separately in sections 2.3─2.5. Table 2.2 gives the comparison of some typical methods in the three groups, by considering the datasets used, the number of target density categories, classifier, and feature selection procedure, from which we can see the following gaps in research:

1) Multi-density (3 or 4 categories) classification is a challenging work, with only a few reported results surpassing 80% accuracy.
2) Recent studies based on texture analysis have shown promising classification performance with over 80% accuracy, but these methods have the problems of high feature dimensionality and lack of spatial information.
3) Feature selection was not considered and analysed carefully in related work. Some work skipped this step or others did not present a comparative analysis.
4) The performance of density estimation models can be influenced by different factors: the accuracy of ground truth labels (i.e. disagreement between readers), the quality of original images (i.e. FFDM or SFM), target number of categories (i.e. classification criteria), experimental test methods.

Based on the literature review in Section 2.3─2.5 and the summary in Table 2.2, four classification models are developed and introduced in **Chapter 5─7**, with different considerations of image feature extraction methods. Two mammogram datasets, MIAS and INbreast (introduced in **Chapter 3**), are used to test the proposed models, and experimental results are evaluated and analysed using multiple metrics.

Table 2.1: A summary of representative methods for MCs detection with test datasets, results evaluation methods and detection performance.

References	Methods	Dataset / Number of images tested	Research aim / Experimental results
(Guo et al., 2016)	Non-linking simplified pulse coupled neural network method	MIAS dataset JSMIT dataset Private dataset	Aim: Detecting MC clusters Results: Specificity = 94.7%, Sensitivity = 96.3%, AUC = 0.92
(Mehdi et al., 2017)	Automatic non-linear stretching scheme and Shannon entropy based wavelet coefficient thresholding	Mini MIAS dataset	Aim: Detecting MC clusters Results: (FROC curves) Average number of FP clusters per image = 0.48 when sensitivity = 97%
(Liu, Mei, Liu & Hu, 2015)	Possibilistic fuzzy c-means clustering algorithm and weighted support vector machine	INbreast dataset	Aim: Detecting MC clusters (tiny MCs are ignored) Results: (FROC curves) False positive rate = 2.3 clusters/image when sensitivity = 92%
(Alasadi & Al-Saedi, 2017)	Two dimensional discrete wavelets transform	MIAS dataset, 65 images were used	Aim: MCs cluster detection Results: Sensitivity = 93.1%
(Cea, Yang, & Nishikawa, 2018)	MCs' spatial distribution patterns and image features related to individual MCs	Private dataset, 186 images were used	Aim: Reduce false positives in detected MCs, classifying MCs clusters Results: (ROC curves) AUC = 0.690 at sensitivity = 63%, Average fraction of false positives is 16% at sensitivity of 50%.
(de Cea, Nishikawa, & Yang, 2018)	Mahalanobis distance (MD) outlier detector and Stochastic Neighbour Graph (SNG) outlier detection method	Private dataset, 188 images were used	Aim: MCs detection Results: (FROC curves) FP number per unit area = 1.43 ± 1.34 when sensitivity = 75% (based on an SVM detector and SNG)
(Wang, Yang, & Nishikawa, 2013)	A linear structure detection procedure with a dual-thresholding scheme	Private dataset 200 images were used	Aim: FPs reduction and MC clusters detection Results: (FROC curves) FPs were reduced by 30%, average number of FP clusters = 1.05 when sensitivity = 85%
(Domingues & Cardoso, 2014)	Bayesian Surprise method	INbreast dataset	Aim: MCs detection Results: (FROC curves) Average FP number per image = 108 when sensitivity = 60.3%
(Wang & Yang, 2018)	Context-sensitive deep neural network	Private dataset 292 images were used for testing	Aim: MCs detection Results: (FROC curves) FP rate = 1.03 FPs/cm^2 when sensitivity = 80%

Table 2.2: A summary of representative methods for MD evaluation with testing datasets, results evaluation, classification performance, and the choice of classifier and feature selection methods.

	References	Methods	CA (%), dataset (categories)	Classifier	Feature selection
Heterogeneous features	(Muhimmah & Zwiggelaar, 2006)	Multi-resolution Histogram features	77.57 on MIAS (3)	DAG-SVM	No feature selection step
	(Subashini, Ramalingam, & Palanivel, 2010)	Statistical features including mean, standard deviation, third moment, etc.	95.44 on MIAS (3) (only 43 images tested)	SVM	No feature selection step
	(Chen, Denton, & Zwiggelaar, 2011)	Image-patch texton	75.0 on MIAS (4)	KNN	No feature selection step
	(Tzikopoulos, Mavroforakis, Georgiou, Dimitropoulos, & Theodoridis, 2011)	21 features combined intensity statistical features and fractal dimension related features.	77.02 on MIAS (3)	CART, KNN, SVM	No feature selection step
	(Mario, Mislav, & Krešimir, 2012)	GLCM + intensity features	82.0% on MIAS (3) 79.2% on MIAS (4)	KNN	Wrappers
	(Oliver et al., 2015)	Combined intensity, texture and morphological features	Dense tissue segmentation	SVM	No feature selection step
	(George, Rampun, Denton, & Zwiggelaar, 2016)	Multi-scale blob detection	No CA results	No classifier	No feature selection step
	(Qu et al., 2020)	GLCM features	94.04% (4)	RF	No feature selection step
DL based features	(Kallenberg et al., 2016)	Deep learning	Dense tissue segmentation	–	–
	(Lee & Nishikawa, 2018)	Deep learning	Dense tissue segmentation	–	–
	(Mohamed et al., 2018)	Deep learning	AUC=0.98 on private dataset (2)	–	–
	(Li et al., 2018)	Deep learning / feature based statistical learning	Dense tissue segmentation	–	–
Texture descriptors	(Rampun, Morrow, Scotney, & Winder, 2017)	LTP	82.33 on MIAS (4)	SVM	No feature selection step
	(Rampun, Scotney, Morrow, Wang, & Winder, 2018)	LQP with multi- scales, topologies, orientations	86.13 on MIAS (4) 82.02 on INBreast (4)	SVM	Dominant feature patterns
	(George & Zwiggelaar, 2019)	LBP variants	77.38 ± 1.06 on MIAS (3)	Bayesian network	Correlation based feature subset (CFS) selection method
	(Rampun, Morrow, Scotney, & Wang, 2020)	LSP	83.3 ± 8.8 on MIAS (4) 77.1 ± 10.9 on INbreast (4)	SVM, RF, MLP, KNN	Dominant feature patterns

3 MATERIALS AND METHODS

This chapter introduces the mammogram datasets used in this thesis for testing the proposed methods. In addition, we outline the pre-processing steps and all the main methods used in the developed models and discuss the metrics adopted in this thesis for evaluating the experimental results.

3.1 Materials

Two mammogram datasets, INbreast (Moreira et al., 2012) and MIAS (J Suckling et al., 2015), are used in this study to test the proposed MCs detection framework and different MD classification models.

INbreast (Moreira et al., 2012) is a full field digital mammograms (FFDM) dataset and consists of 409 images from 115 cases, including bilateral mediolateral oblique (MLO) and craniocaudal (CC) views. Each image is saved in the DICOM format and in sizes of 3328 × 4084 or 2560 × 3328 pixels, depending on the compression plate used in the acquisition. This dataset offers carefully associated ground truth (GT) annotations made by a specialist in the field and validated by a second specialist. There are 244 images in INbreast containing calcifications, in which pixel level annotations of individual MC spots are manually identified by radiologists as ground truth. Fig. 3.1 illustrates some examples of INbreast mammograms with MCs. For breast density classification, the density class of each mammogram is labelled as one out of four BI-RADS categories (BI-RADS I-IV). The distribution of density labels in the dataset is as follows: 136 (BI-RADS I or fatty), 146 (BI-RADS II or scattered density), 99 (BI-RADS III or heterogeneously dense) and 28 (BI-RADS IV or extremely dense).

Examples of INbreast mammograms in different density categories are shown in Fig. 3.2.

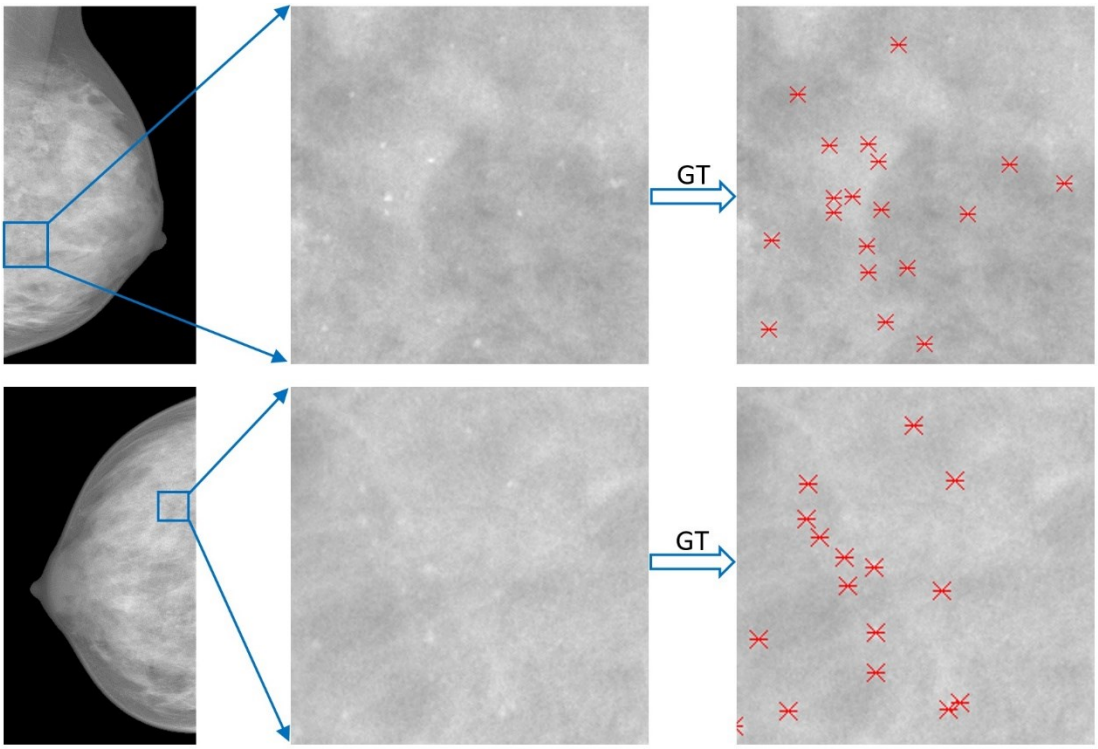


Figure 3.1: Two examples of INbreast mammograms with microcalcifications, and ground truth for individual microcalcification spots identified by radiologists and marked using red stars in the third column.

MIAS (J Suckling et al., 2015) is a scan field mammograms (SFM) dataset, containing 322 images (161 women) with only MLO views on both sides from the UK National Breast Screening Programme. Each mammogram is at 50 micron resolution in "Portable Gray Map" (PGM) format. All images are associated with tissue density ground truth labels of three classes: fatty (F), fatty-glandular (G) or dense-glandular (D). There are 106 images belonging to the fatty group, 104 and 112 images to the fatty-glandular and dense-glandular classes. Fig. 3.2 gives examples of images in the dataset belonging to different density classes.

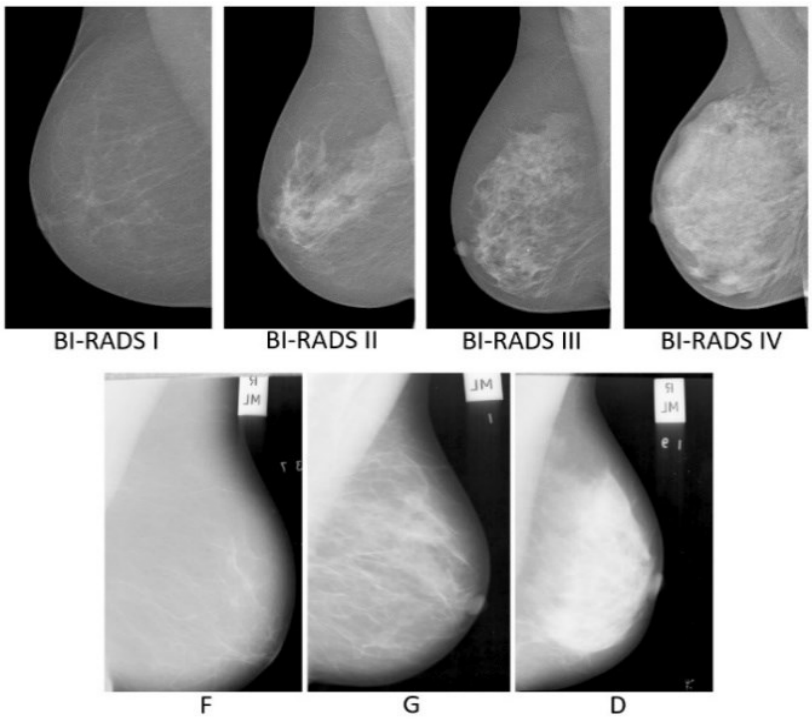


Figure 3.2: First row: INbreast mammograms in four BI-RADS density classes; Second row: MIAS mammograms in three tissue density classes.

3.2 Mammogram Pre-processing

This section outlines the pre-processing steps in both microcalcification detection and breast density estimation tasks.

The main tasks performed in the pre-processing stage are breast area segmentation, denoising operation and resizing the image (for density classification). The background region of scan-field mammograms usually contains a label (e.g. left or right side marker). In MLO view mammograms, the pectoral muscle region is also captured along with the breast region. However, the pectoral muscle represents a predominantly dense region which may easily affect breast density evaluation. Therefore, breast segmentation is first applied to remove non-breast areas such as the background region, pectoral muscle and label artefacts. A multifractal method (introduced in **Chapter 4**) is used to enhance the contrast between image background and the breast tissue region. An intensity thresholding method and morphological operations (Mustra, Grgic, & Rangayyan, 2016) are used to separate the breast region and artefacts from the background. The artefact components can be recognized and then removed by keeping only the largest connected area (breast region). A K-means algorithm and polynomial fitting approach (Slavković-Ilić, Gavrovska, Milivojević, Reljin, & Reljin, 2016) are

employed to distinguish the pectoral muscle region from the breast region in MLO view mammograms. A median filter of 3 × 3 size is used to reduce noise. Finally mask images are obtained, which are used to extract image features from only the region of interest (breast area) in the following steps. Fig. 3.3 and Fig. 3.4 show some examples of segmenting the breast region from image background using INbreast and MIAS mammograms respectively.

However, the datasets contain a few images with blurred boundary lines between the pectoral muscle area and the breast region, which result in inaccurate mask images. There are approximately 20% mammogram images in the two datasets used in this study belonging to this kind of challenging case, and manual operations are used in these mammograms to guarantee outputting proper mask images for the following experiments. Fig. 3.5 shows examples of challenging cases and their mask images before and after using manual operations.

For the MIAS dataset, which contains only MLO view images, related work (George & Zwiggelaar, 2019) (George, Denton, & Zwiggelaar, 2018) demonstrated that using a cropped square region of interest (ROI) brings a better classification result. This study therefore uses the method in (George, Denton, & Zwiggelaar, 2018) to extract the ROIs in MIAS. Fig. 3.6 illustrates ROI extraction in MIAS mammograms.

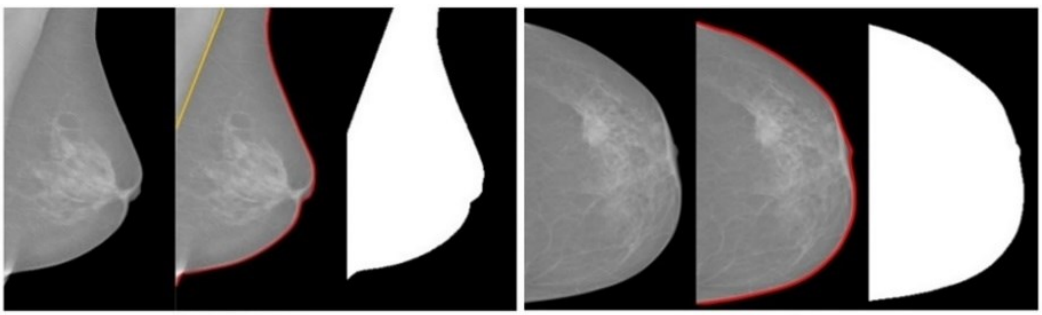


Figure 3.3: Two examples from the INbreast dataset, showing the processes of breast region segmentation and mask generation

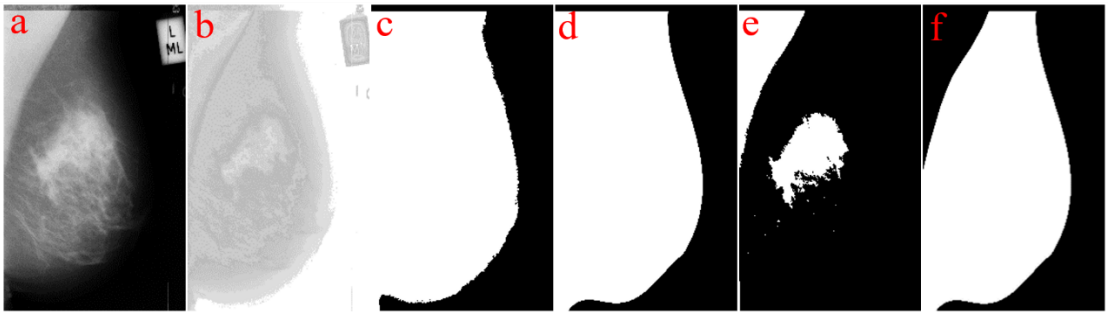


Figure 3.4: Segmentation of breast region in MIAS mammograms. (a) Input mammogram; (b) Enhancing breast region and artefact areas; (c) Rough contour of breast region; (d) Smoothing breast contour; (e) Finding pectoral muscle area; (f) Breast mask image without pectoral muscle region.

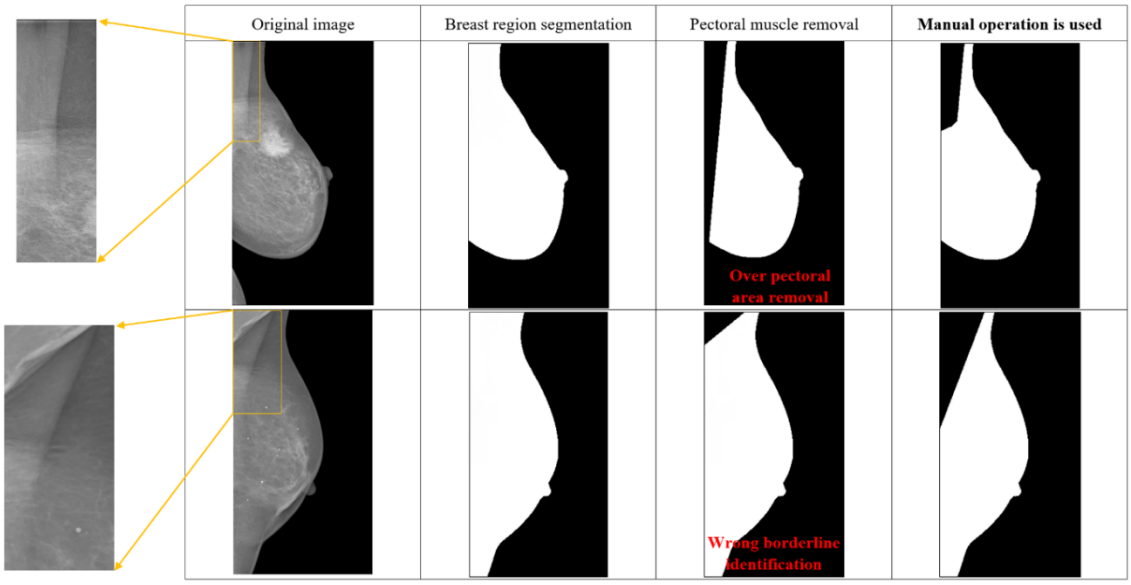


Figure 3.5: Challenging cases with inaccurate mask images generated and their adjusted mask images based on manual operations.

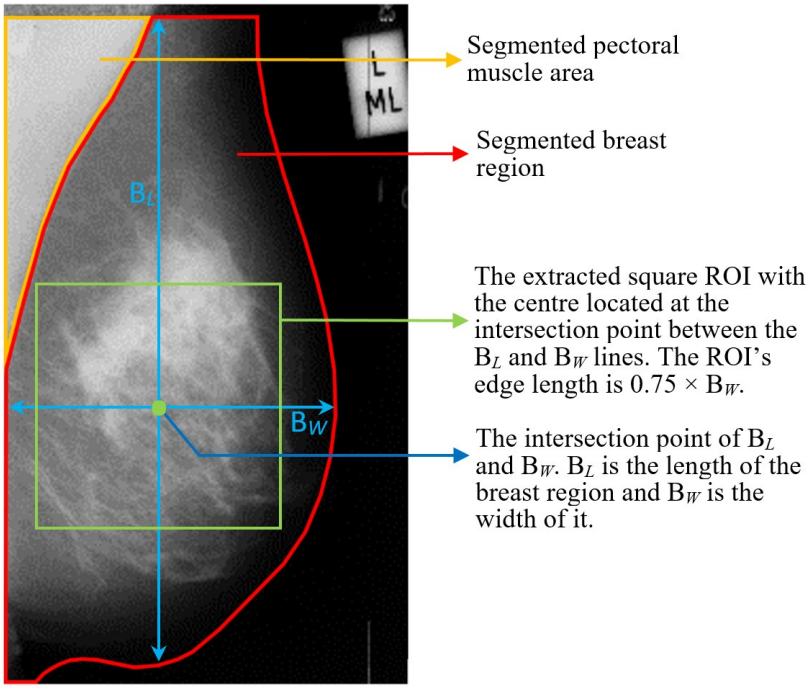


Figure 3.6: ROI extraction from the central area of MLO view mammogram in MIAS dataset.

3.3 Methods

The work in this thesis considers and adopts a series of methods to develop MCs detection and MD classification models based on mammogram analysis. Table 3.1 lists the main methods used for the above tasks. These methods are employed and improved with different research aims, such as ROI segmentation, enhancing image contrast, feature extraction, feature selection, etc. More details of applying these methods and improvement strategies developed in this study are given in the following chapters.

Table 3.1: Dominant methods and datasets used in this thesis and the number of images used to test the proposed models in experiments.

	Methods	Dataset and the number of images used
Pre-processing (Chapter 3)	1. Multifractal analysis 2. Intensity thresholding 3. Morphological operations 4. Polynomial fitting 5. Median filter	INbreast (409 images) MIAS (322 images)
MCs detection (Chapter 4)	1. Linear structure detector 2. Weber's law based multifractal measure 3. Support vector machine 4. Patch-wise CNN	INbreast (244 images containing MCs) 9444 patches
MD classification model 1 & 2 (Chapter 5)	1. Multifractal spectrum 2. Histogram shape analysis 3. Chi-square test 4. Local binary patterns 5. Autoencoder network	INbreast (78 images are tested in model 1 and 409 images are tested in model 2)
MD classification model 3 (Chapter 6)	A novel texture descriptor: rotation invariant uniform local quinary patterns (RIU4-LQP)	INbreast (409 images) MIAS (322 images)
MD classification model 4 (Chapter 7)	1. RIU4-LQP based histogram 2. Baddeley's K-inhom function 3. Recursive feature elimination	INbreast (409 images) MIAS (322 images)

3.4 Evaluation Metrics

For the two research tasks, this thesis adopts multiple measurements to assess the experimental results and conducts quantitative analysis. To evaluate the MCs detection results, this thesis uses a confusion matrix and free-response receiver operating characteristic (FROC) method; in the breast density classification task, this thesis uses classification accuracy, area under the receiver operating characteristic (AUC) value, Kappa coefficient and a statistical test (t-test) to assess the classification results. These metrics are discussed in brief below.

Confusion Matrix and Related Metrics

A confusion matrix is a table with rows and columns that report the number of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values.

Based on this table, other parameters can be calculated to give a more detailed analysis of the experimental results. Fig. 3.7 shows an example of a confusion matrix for binary classification.

		Predicted class	
		P	**N**
Actual class	**P**	**TP**	FN
	N	FP	**TN**

P = Positive; N = Negative.

Figure 3.7: An example of confusion matrix with two classes.

Other evaluation metrics such as sensitivity or true positive rate (TPR), specificity or true negative rate (TNR), accuracy, F1 score and Kappa coefficient (Cohen, 1960) (Artstein & Poesio, 2008) are also used in this thesis. They are defined as follows.

$$Sensitivity = \frac{TP}{TP + FN} \tag{3.1}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3.2}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.3}$$

$$F_1 = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3.4}$$

$$Kappa = \frac{Accuracy - RandomAccuracy}{1 - RandomAccuracy} \tag{3.5}$$

$$RandomAccuracy = \frac{(TN + FP) \times (TN + FN) + (FN + TP) \times (FP + TP)}{Accuracy \times Accuracy} \tag{3.6}$$

For the first research task (i.e. MCs detection), the developed methods were evaluated using sensitivity calculated as the number of correct MC predictions divided by the total number of MC spots (i.e. ground truth), and the specificity calculated as the number of correct non-MC predictions divided by the total number of non-MC spots (i.e. the total pixel number – ground truth number). In addition, the FP number in the confusion matrix is analysed separately, due to the concern of the high FP number

affecting the overall detection accuracy negatively as discussed in Chapter 2. Therefore, by using the metrics of sensitivity, specificity and the FP number, we can produce more informative quantitative analysis in the evaluation of the proposed detection methods including the non-MC detection rate and the corresponding FP number. The experimental results are presented in Chapter 4.

For the second research task (i.e. breast density classification), in addition to using the accuracy to assess the classification results, the F1-score and the Kappa coefficient are also considered in this study. F1-score becomes high when both precision and recall are high. F1-score is the harmonic mean of precision and recall and is a better measure than accuracy. The Kappa coefficient takes imbalance in class distribution into account and can provide supplementary information in the analysis of classification performance. As the breast density estimation in this study is a multi-category classification task (3 or 4 density categories) and two datasets containing unequal distribution of image numbers between different categories are used (as discussed in Chapter 3), Kappa coefficient is found to be a useful metric. Our experimental results of breast density classification are presented by using the classification accuracy, F1-score, and Kappa coefficient, in Chapters 5-7.

Area under the Receiver Operating Characteristics (AUC)

The ROC curve shows the performance of a classification model by plotting sensitivity-vs-(1 − specificity) at different classification thresholds. AUC measures the entire two-dimensional area underneath the ROC curve from (0, 0) to (1, 1). AUC ranges in value from 0 to 1 and a model whose predictions are 100% correct has an AUC of 1.0, and vice versa. In Chapters 5─7, this thesis presents four different breast density classification models and AUC values are calculated to evaluate the classification performance.

Free-Response Receiver Operating Characteristic (FROC)

FROC curve is a plot of the sensitivity and the number of false detections per image (Samuelson & Petrick, 2006). Comparing to the commonly used ROC curve, the FROC method additionally requires the observer/detection method to locate abnormalities, which is suitable for the MCs detection work, as it aims to detect positions of MC spots. In Chapter 4, we generate plots of FROC curves to analyse the experimental results and to compare with other methods as well.

3.5 Chapter Summary

This chapter introduces the two datasets used for testing the proposed methods in this thesis with different research aims. Both datasets are publicly available and have ground truth information offered by clinicians. Experimental results in this thesis are compared with related work in the literature based on the above test sets. In addition, this chapter presents pre-processing methods used in the detection of microcalcifications and analysis of breast density and outlines important evaluation metrics used. An overview of methods employed in different parts of this thesis is given and more details are elucidated in the subsequent chapters.

4 MICROCALCIFICATION DETECTION USING TEXTURE FEATURES AND CNN

This chapter presents a complete framework of microcalcifications (MCs) detection, which contains two parts: MCs detection (stage 1) and false positive (FP) reduction (stage 2). An overview of the framework is given in Fig. 4.1. The aim of the MCs detection task in this thesis is to automatically detect individual MC spots and locate their positions in mammograms, offering the detection results to radiologists for aiding breast cancer diagnosis. A MCs detector is first developed by using a linear structure operator and multifractal features to produce the initial detection results. Based on the detected MCs set, a convolutional neural network (CNN) classifier is proposed to reduce the FP number and improve the overall detection accuracy. A digital mammogram dataset INbreast is used to test the proposed detection framework. Experimental results in the two stages are analysed separately and compared to other methods in the literature. Comparative analysis between existing methods demonstrates the superiority of the proposed MCs detection framework.

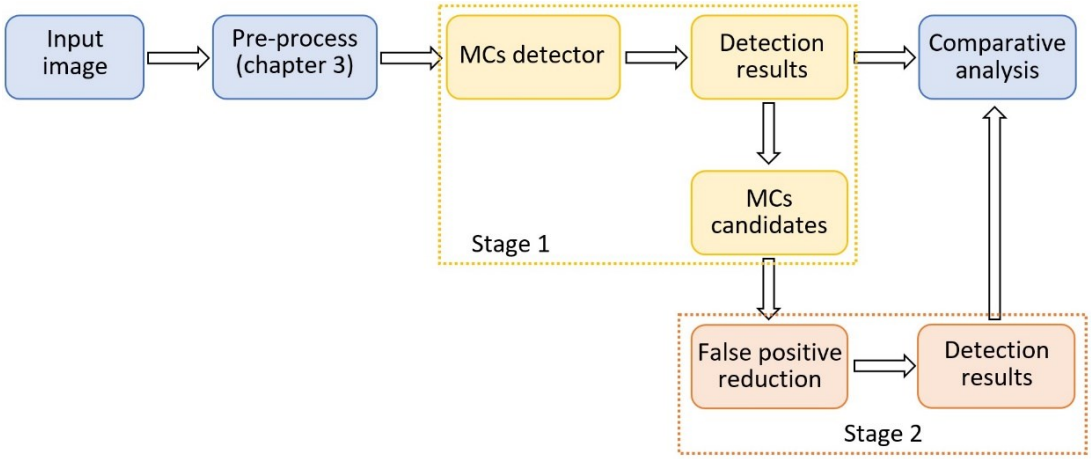


Figure 4.1: An overview of the proposed MCs detection framework with two different stages.

4.1 MCs Detection in Mammograms

MCs' size, shape and local contrast can have large variations in each mammogram. This variation could lead to false positive outcomes affecting detection accuracy. According to radiologists, the size of a single MC spot can have a size between 0.1 and 1 millimetre. For example, if the image resolution of mammograms is 50 μm per pixel, then a MC spot with diameter size 0.1mm corresponds to only two pixels, and 0.5mm corresponds to ten pixels. If there are over 3 MC spots within a 1 cm^2 area, we need to further consider clusters of MCs and their benign or malignant category. Accurate MC spots detection in mammograms is a challenging task, as small MC spots have low contrast relative to other components, such as glandular or fibrous tissue, which could be recognized as MCs incorrectly by detection algorithms. In addition, the detection accuracy of individual MCs also impacts subsequent detection performance of MC clusters which relates to the number and distribution of MC spots in a small local area (Cea, Nishikawa, & Yang, 2017).

4.2 MCs Detector

A pipeline containing the main steps in the proposed MCs detector is shown in Fig. 4.2. The breast region is segmented first in each original mammogram in the pre-processing step (chapter 3). Based on the characteristics of the MCs as outlined in Section 4.1, we crop local regions into small patches with two different sizes: 32 × 32 pixels and 128 × 128 pixels for detecting MCs in different sizes. To avoid incorrectly recognizing glandular or fibrous tissue as MCs and increasing the FP number, a linear

structure detector is employed to differentiate line structures from MCs. Multifractal analysis is used to extract MC-related texture features for distinguishing MCs in the background of breast tissue. More details of the proposed detector are given in the following sections.

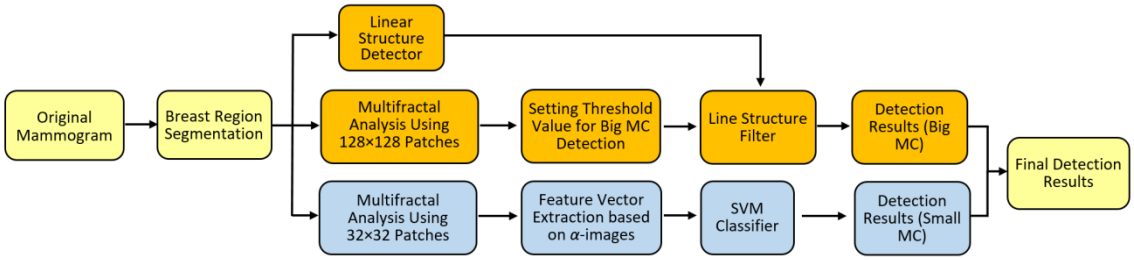


Figure 4.2: The pipeline of the proposed MCs detector.

4.2.1 Multifractal Analysis

Multifractal analysis, which depends on the choice of local measures, can be used to describe image texture features that are useful in classification tasks (Ibrahim & Mukundan, 2015) (Paskas, Reljin, & Reljin, 2016) (Xue & Bogdan, 2017). Let $\mu_w(p)$ denote a multifractal measure function, where p is the central pixel in a square window of size $w \times w$. Then, a local singularity coefficient, Hölder exponent or α-value (Falconer, 2005), can be calculated to reveal variation of the selected $\mu_w(p)$ function within the neighbourhoods of the pixel p.

$$\mu_w(p) = Cw^{\alpha_p}, w = 2i+1,\ i = 0,1,2...,d \tag{4.1}$$

$$\log(\mu_w(p)) = \alpha_p \log(w) + \log(C) \tag{4.2}$$

where, C is an arbitrary constant and d is the total number of windows used in the computation of α_p. The value of α_p can be estimated from the slope of a linear regression line in a log-log plot where $log(\mu_w(p))$ is plotted against $log(w)$. Commonly used multifractal measures for calculating α are outlined below:

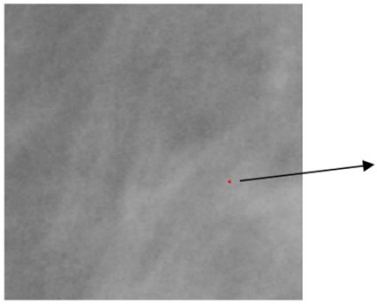
$$Maximum: \ \mu_w(p) = \max_{(k,l)\in\Omega} g(k,l) \tag{4.3}$$

$$Inverse-Minimum: \ \mu_w(p) = 1 - \min_{(k,l)\in\Omega} g(k,l) \tag{4.4}$$

$$Summation: \ \mu_w(p) = \sum_{(k,l)\in\Omega} g(k,l) \tag{4.5}$$

$$Iso: \ \mu_w(p) = \#\{(k,l) | g(p) \cong g(k,l), (k,l) \in \Omega\} \tag{4.6}$$

where, $g(k, l)$ represents the intensity value of a pixel at position (k, l); Ω denotes the set of all neighbourhood pixels of p in the window; # is the number of pixels in a set. Pixel intensity values are normalized into the range of [0, 1] when considering Maximum and Inverse-Minimum measures. Such normalization brings better image enhancing results when computing the Hölder exponent due to the amplifying effect of the logarithmic function. A patch of one mammogram is shown in Fig. 4.3 and one pixel p (marked in red colour) is chosen for illustrating the calculation of the α value. Fig. 4.4 shows the measured values of $\mu_w(p)$ by using maximum measure when the square window size w is 1, 3, and 5 respectively. The α value can be estimated using the slope of the linear regression line in a log-log plot.



148	148	150	152	155
152	156	151	147	153
151	154	151	150	157
149	156	152	148	155
155	159	159	155	155

Figure 4.3: (a) A local region (200 × 200 pixels) in one mammogram, and (b) the central pixel p(151) and its neighbourhood.

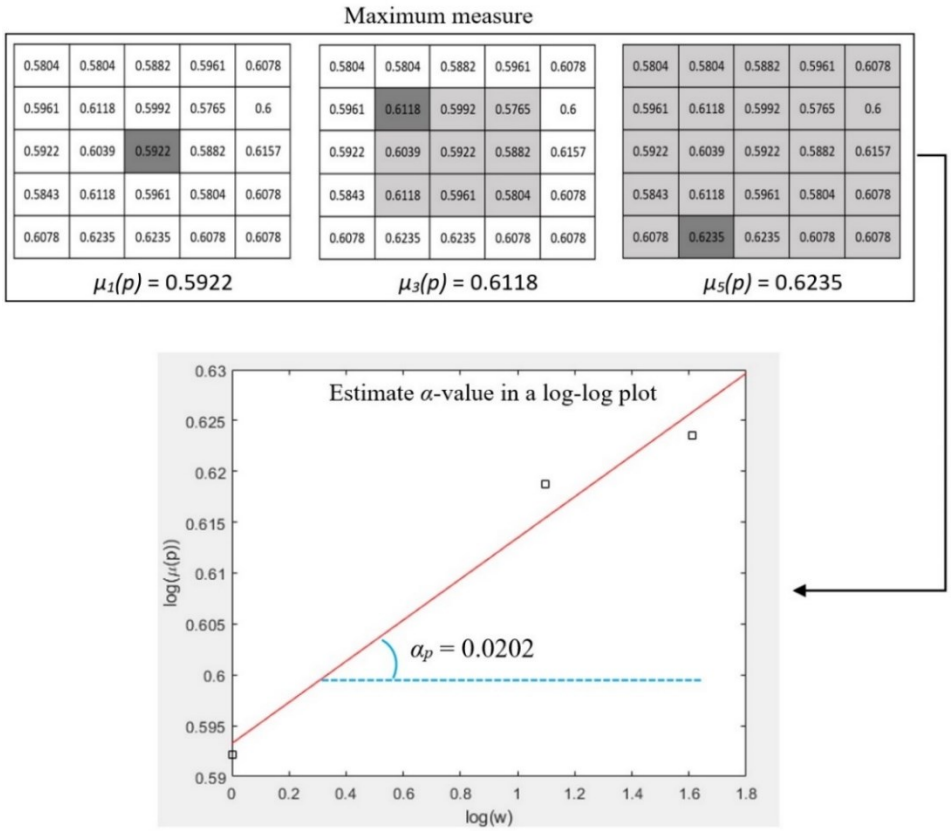


Figure 4.4: An example of estimating α value for the point p in Fig. 4.3 using the maximum measure.

4.2.2 Alpha Image and Texture Enhancement

Alpha images are obtained using α_p to replace the intensity value at each position p. In α-images, certain texture features and patterns have significantly higher contrast compared to original images. The range of α values in an α-image is denoted by [α_{min}, α_{max}]. This range is subdivided into a set of bins, and pixels having the α values in the same bin are counted to obtain an α-histogram which can be further used as texture features (Ibrahim & Mukundan, 2014) (Ibrahim & Mukundan, 2015) (Reljin, Reljin, Pavlovic, & Rakocevic, 2000). Fig. 4.5 shows examples of α-images and their α-histograms using four multifractal measures.

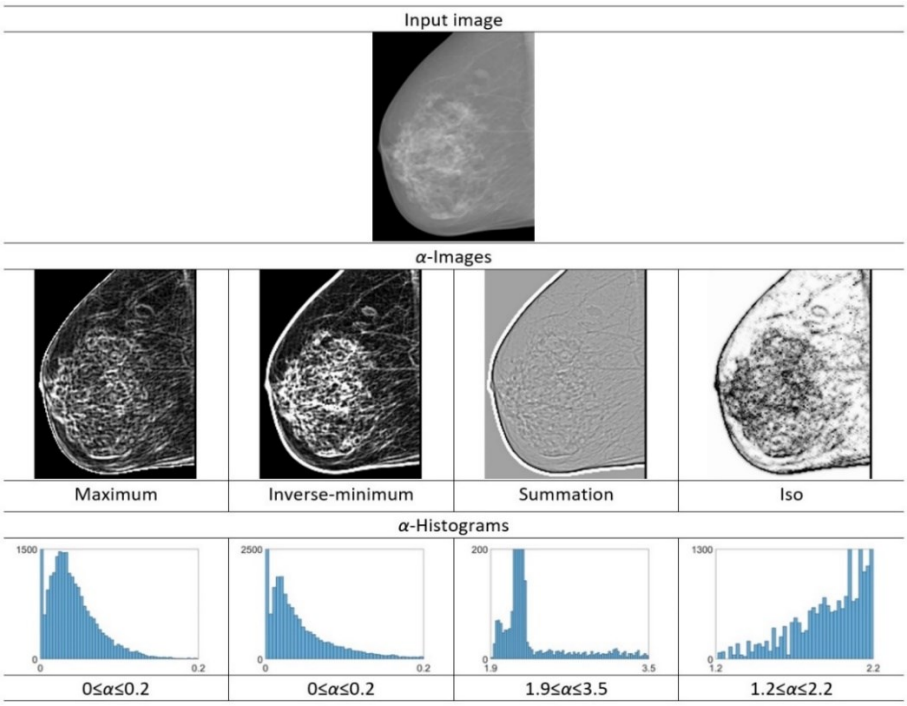


Figure 4.5: Examples of α-images and corresponding α-histograms using different multi-fractal measures.

For better highlighting texture patterns related to MCs in α-images, the α-range [α_{min}, α_{max}] could be further subdivided into some subintervals with narrower α-ranges. In each subinterval [α_i, α_{i+1}], only the pixels possessing α values in this range are retained, thus effectively enhancing specific image features. An α-image only containing the pixels in one subinterval is called an α-slice. In our experiments, we find that some α-slices enhance texture features significantly in mammograms, which could be used to identify the ROI and extract relevant features. In this research work, such characteristics are used to detect MCs from breast tissue background. As seen in Fig. 4.6, α-slices with α values in a narrow range help in enhancing texture patterns related to MC spots.

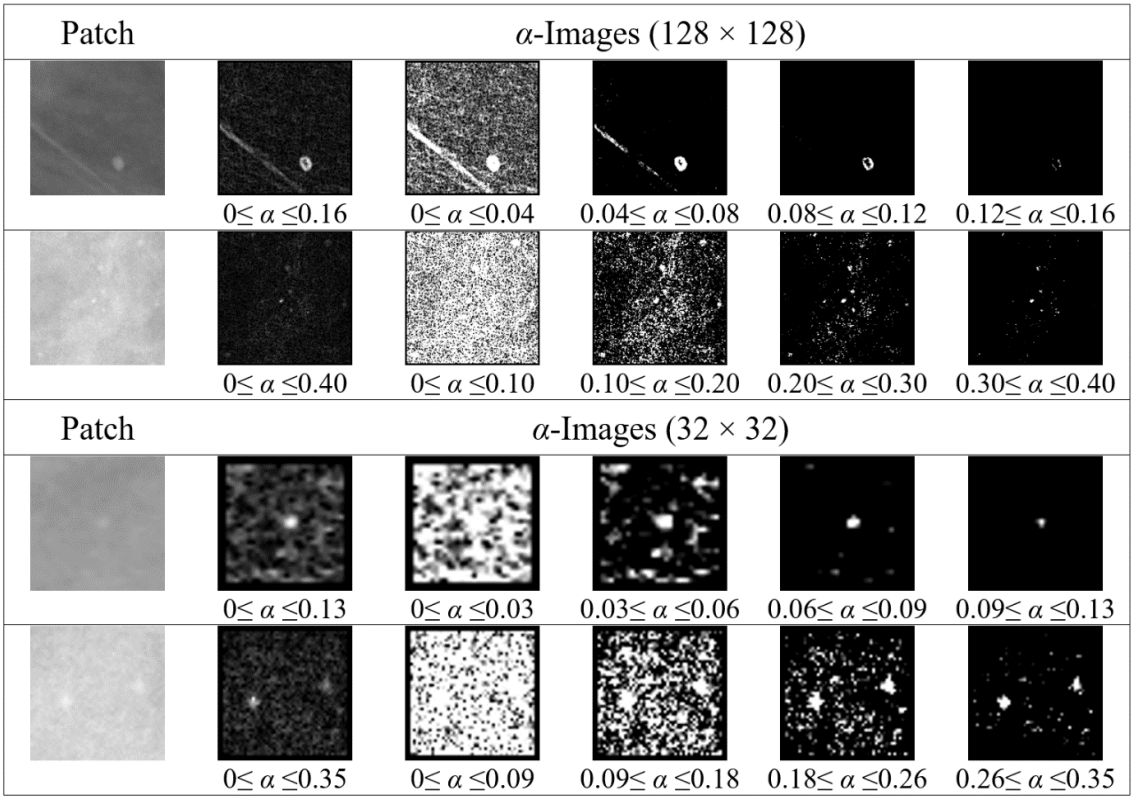


Figure 4.6: Sub-images with MC spots contained and their *α*-slices (using Inv-min measure) in different *α* value ranges.

4.2.3 Alpha Value Range Selection

Since the pixel intensity and the local tissue density vary differently among different mammograms or local regions, the α-range $[\alpha_{min}, \alpha_{max}]$ keeps changing with images and multifractal measures used. It is difficult to select one single α-value range in which the α-slice best describes MC features. In our experiments, we found that MCs textures could be highlighted in a higher α-value sub-range better than that in a lower sub-range, as illustrated in Fig. 4.6. This can be ascribed to the slightly clearer local contrast of MCs in small local region areas, which generates higher α values. In the proposed MCs detector, adaptive α-value range selection rules are designed for 128 × 128 and 32 × 32 sized sub-images separately as follows. For one sub-image, a specific percentage (PA) of total points is used to select a suitable α-value range, which means that there exists an α threshold value denoted by α_t, and it satisfies:

$$\sum_{i=t}^{\max} n(\alpha_i) = PA \times Total\ pixel\ number \tag{4.7}$$

where, $n(\alpha_i)$ denotes the number of points possessing the α_i value in this α-image. In our experiments, different values of PA (from 0.01 to 0.1) are tested, and PA = 0.04 is used

to enhance MCs patterns in 128 × 128 sized sub-images and $PA = 0.1$ for 32 × 32 sized sub-images.

4.2.4 Linear Structure Detector

Glandular and fibrous tissue regions having a nearly linear structure and high local contrast could potentially be misclassified as MCs. Therefore, a linear structure detector algorithm is used in the proposed method for identifying such tissue structures. Research work in (Wang, Yang, & Nishikawa, 2013) demonstrated its effectiveness in detecting linear structures in mammograms. Traditionally, the linear structure detector is used as follows (Zwiggelaar, Astley, Boggis, & Taylor, 2004).

$$S(x) = L(x) - N(x) \tag{4.8}$$

$$L(x) = \max_{\theta_i \in [0,\pi)} L_{\theta_i}(x) \tag{4.9}$$

where, x denotes the location information of pixels in mammograms; $S(x)$ is the line strength signal; $N(x)$ denotes the average local background intensity around x; and $L_{\theta_i}(x)$ is the average grey-level in the orientation of θ_i. In our experiments, θ_i uses 12 equally-spaced orientation angles. Here, a linear structure is defined as a straight line in length of at least 25 pixels and with a specific angle of θ_i, keeping the pixel x in its middle point. A 5 × 5 square window with x in the centre is considered as the local background area when computing $N(x)$. Comparing to commonly used line detectors, our proposed method does not use pixel intensity values to calculate $L_{\theta_i}(x)$ and $N(x)$ but uses each pixel's α value to measure $S(x)$. Fig. 4.7 shows an example of using the linear structure detector.

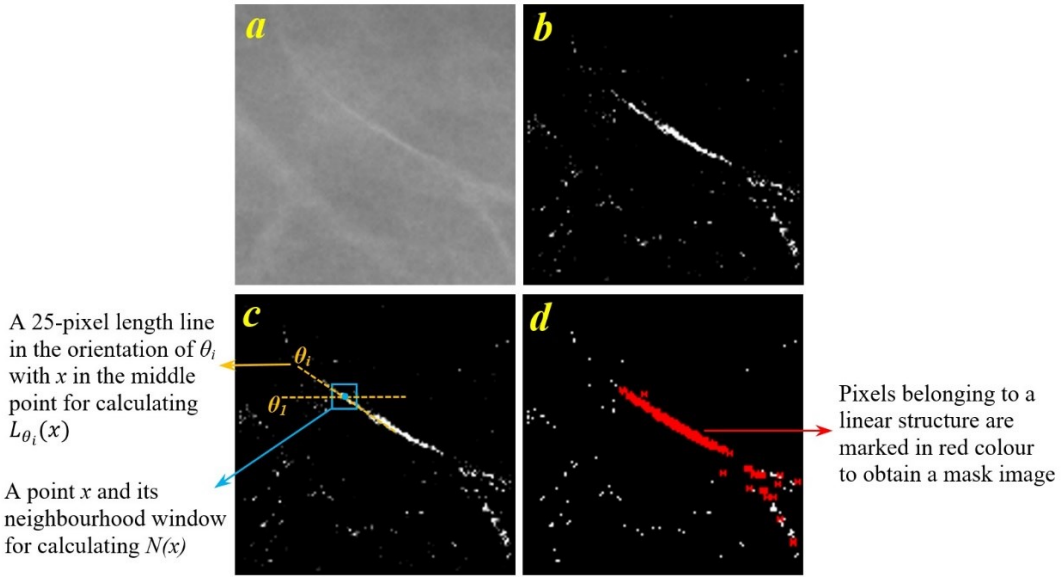


Figure 4.7: (a) An original mammogram patch containing fibro-glandular tissue; (b) Texture enhancement in an α-slice; (c) Applying the linear structure detector by using α values (i.e. in α-slice); (d) Linear structure detection result.

4.2.5 MC Spots Detection

Due to the heterogeneous features of MCs, it is difficult to detect individual MC spots by using identical rules. The diameter of MC spots can range from 0.1 mm to over 0.5 mm. By considering this fact, two scales of sub-images, 32 × 32 pixels and 128 × 128 pixels sliding windows, are proposed to detect varied sized MCs.

Big MCs detection. Specifically, a threshold value T_{area} for defining the size of big MCs is needed, and T_{area} = 25 pixels is assigned in our experiments, aiming at recognizing potential MC spots with area greater than 25 pixels or diameter greater than 5-pixel length in 128 × 128 sized sub-images. Since the shapes of MCs are irregular, for example, they are not limited to round shapes but also could be in oval, rod, stellate, or aciform shapes. Therefore, there are no morphological detection rules designed for discerning MCs in this method. When there are too many overlapping points (T_{over}) between a detected MC spot and a linear structure, the current spot will not be considered as a MC, as the detected object possibly is part of a fibro-glandular tissue.

Small MCs detection. For small MC spots with area less than 25 pixels or diameter less than 5 pixels, 32 × 32 sized patches are used to detect the existence of small MCs. Some tiny MCs occupy only 2 to 4 pixels in area and do not possess as high image contrast as other big MCs or glandular tissue; therefore, it is almost impossible to detect

such MC spots by using global texture features. However, after narrowing the sliding window size to 32 × 32 pixels and performing the multifractal based texture enhancing scheme (Section 4.2.2−4.2.3), those tiny MC spots are highlighted significantly in this local region and can be distinguished from the tissue background.

SVM classifier. An SVM classifier based on α-values and intensity information is trained and used with the aim of detecting tiny MCs in each patch. A feature vector X used in this classifier consists of six features: $X(p) = [\alpha_1, \alpha_2, \alpha_3, i_1, i_2, i_3]$, where p is the currently considered pixel, and $\alpha_1, \alpha_2, \alpha_3$ are the means of α-values calculated from neighbourhood areas of 3 × 3 pixels, 5 × 5 pixels and 7 × 7 pixels respectively around p, and i_1, i_2, i_3 are computed in the same way by using pixel intensity values.

4.3 False Positive Reduction

False positives (FPs) are a common occurrence in the outputs of MCs detectors due to the heterogeneous properties and diverse composition of breast tissue. This section focuses on FP reduction and improvement of the final MCs detection accuracy in mammograms. For further highlighting texture features, a Weber's law based approach is proposed and used to construct a new multifractal measure and the corresponding alpha patches. In order to distinguish MC spots from the candidate set produced by the MCs detector in Section 4.2, a convolutional neural network (CNN) classifier is designed to process original mammogram patches and corresponding alpha patches together for classifying suspicious MC spots to a true positive group or a false positive group. An overview of the MCs detection procedures with the FPs reduction process is shown in Fig. 4.8.

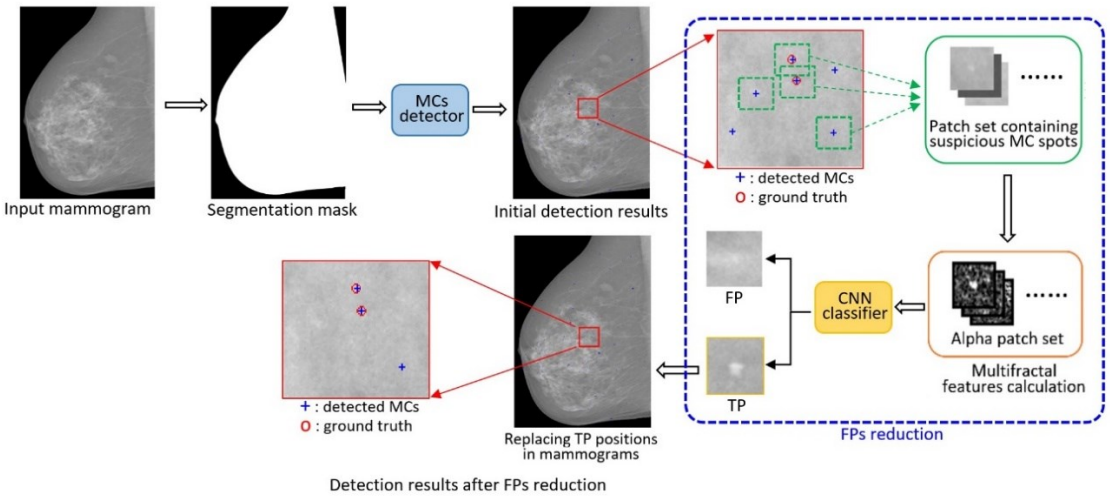


Figure 4.8: An overview of the proposed MCs detection framework with the FPs reduction procedure.

4.3.1 Weber-based Multifractal Local Measure

Among the four multifractal measures introduced in section 4.2.1, the actual intensity value of the central pixel p is not taken into account in the calculations, except when using the Iso measure. In Eqs. (4.3) − (4.5), we can see that the conventional multifractal measures evaluate only neighbourhood pixels information (maximum pixel value, minimum pixel value or the sum of all pixels), ignoring the value at the central pixel p. This violates the aim of the considered task (MCs detection), since we assume that the suspicious MC spot is located in the centre in a small local region, which presents slightly higher local contrast. Changes of intensity between neighbouring pixels and p therefore should be considered to better reflect such local contrast. Weber's law is particularly useful and relevant to this detection task. It states that the perceived change in stimuli is proportional to the initial stimuli (Han, Yen-Wei, & Gang, 2016) (Peng et al., 2017), which can be defined as follows.

$$\frac{\Delta R}{R} = k \tag{4.10}$$

where ΔR is the change of stimuli, R is initial stimuli and k is referred to as the Weber fraction for detecting changes in weight. Inspired by this law, we extend the Maximum measure in Eq. (4.3) to a Weber-based Maximum measure (WMax), which is given by:

$$\frac{\Delta R}{R} = \frac{\max(g(k,l)) - g(p)}{g(p)} = \frac{\max(g(k,l))}{g(p)} - 1 = W - 1 \geq 0 \tag{4.11}$$

$$WMax: \quad \mu_w(p) = \frac{\max_{(k,l)\in\Omega} g(k,l)}{g(p)} = W \geq 1 \tag{4.12}$$

When using the proposed WMax measure with Eqs. (4.11) and (4.12), if p is a totally black coloured point (i.e. $g(p) = 0$), the value of $\mu_w(p)$ will be set to 0 directly to avoid division by zero. The proposed WMax measure takes into account the intensity of the central pixel and enhances the local contrast: a bigger intensity difference leads to a larger computed α value and enhanced texture patterns in a local alpha image. The max measure is selected and extended using the Weber's law, because it is used throughout the following experiments (Section 4.4) with better performance than the other three measures. Fig. 4.9 shows an example of calculating an α value based on the WMax measure, and Fig. 4.10 illustrates an original MC spot-contained patch and its corresponding alpha patches using five local measures.

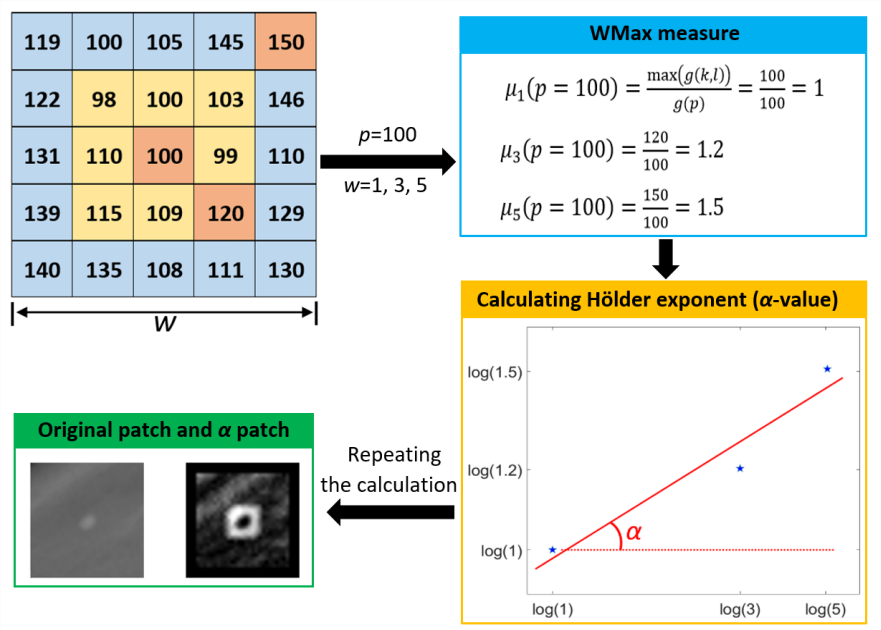


Figure 4.9: An example of calculating an α value and α patch using WMax measure.

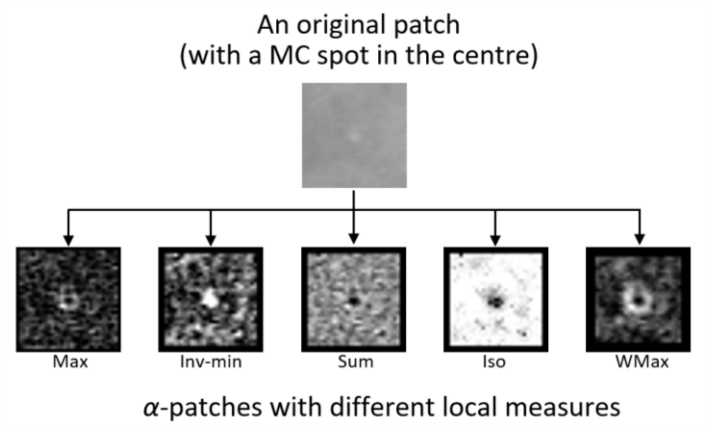


Figure 4.10: An example of one original MC patch and its corresponding alpha patches.

4.3.2 CNN Classifier

This section presents a CNN classifier to automatically learn MCs related image features. A patch-wise CNN architecture is designed to process small local regions which contain suspicious MC spots, with the aim of reducing FPs to a lower level and improving the overall accuracy. Two patch sets, original MC patches and corresponding alpha patches, are fed into the CNN model which can learn image features from both intensity information and multifractal features. As illustrated in Fig. 4.11, two groups of feature maps learned from the MC patch set and alpha patch set respectively are concatenated and then sent to the following fully connected (FC) layers for classifying the input patch into true positive (TP) or false positive (FP). The input patches come from an initial detection result set using a MCs detector which is addressed in Section 4.2.

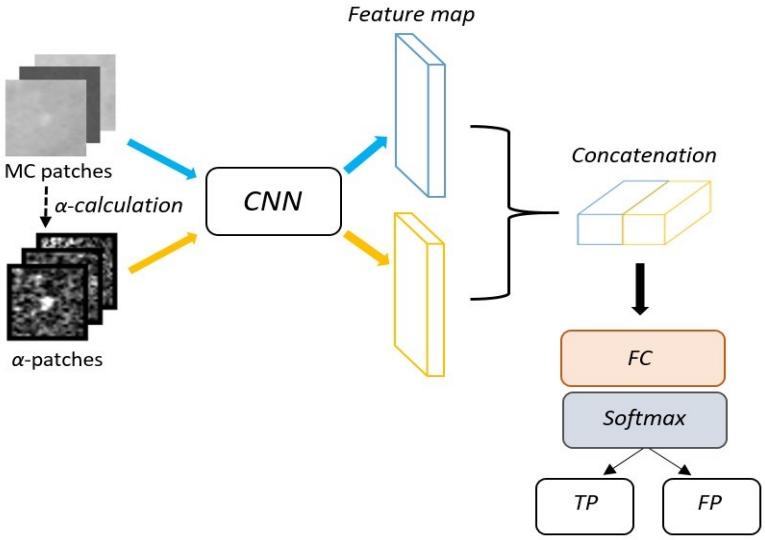


Figure 4.11: The proposed CNN model processing original patch set and alpha patch set separately.

In this proposed CNN model, different numbers of convolutional layers with varying filter numbers are tested in training and cross-validation phases in order to develop a proper model structure. Table 4.1 shows 7 architectures tested in the following experiment. Convolutional layers help to learn MCs features from the input data at different spatial scales. In this study, convolutional kernels are set to be 3 × 3 in size to process input data, then its output is fed into the next layer. A standard rectifier function, rectified linear unit (ReLU) is used as the activation function. Following each convolutional layer, a batch normalization layer is employed to deal with internal covariance shift during the training stage. The output feature maps usually are sensitive to the feature locations in the input, so max-pooling layers are used to down-sample the feature maps, making them more robust to position changes in the image. In addition, convolutional layers can extract image features from increasingly higher scales. A 2 × 2 window with stride 2 is used in each max-pooling layer, which is denoted as 'Pooling' in Table 4.1. Fully connected (FC) layers play the same role as in a feedforward neural network and a softmax activation function is used in the final output layer for giving a binary classification result. Categorical cross-entropy in Eq. (4.13) is used as a loss function.

$$L(y, \hat{y}) = -\sum_{i=0}^{C} (y_i * \log(\hat{y}_i)) \tag{4.13}$$

where, y and ŷ are the ground truth and the predicted value by learning features from local patches and their α-images; C denotes the number of categories.

Table 4.1 Different CNN structures considered in this study. Batch normalization layer after each Conv layer is not listed for brevity. The presence of a layer is marked by '√' sign and the absence by '-' sign.

CNN structure with different filter numbers applied in conv layers							
	#1	#2	#3	#4	#5	#6	#7
Conv	16	20	32	40	64	20	16
Conv	–	–	–	–	–	20	–
Pooling	√	√	√	√	√	√	√
Conv	16	20	32	40	64	32	32
Conv	–	–	–	–	–	32	–
Pooling	√	√	√	√	√	√	√
Conv	–	–	–	–	–	–	64
Pooling	–	–	–	–	–	–	√

4.4 Experiments and Results

This section presents experiments and MCs detection results using the proposed MCs detection framework. The MCs detector and the CNN classifier for reducing FPs are first evaluated separately, and then the final MCs detection results are generated by integrating the two parts. The INbreast dataset containing 244 mammograms with MCs is used to test the proposed methods in this chapter. We use free response receiver operating characteristic (FROC) curve to analyse the experimental results and compare different methods.

4.4.1 Experimental Results Using MCs Detector

As discussed in Section 4.2, the first stage of the proposed framework focuses on developing a MCs detector which uses multifractal analysis, linear structure detector and threshold values. The corresponding parameters and their values used in experiments are listed in Table 4.2.

Table 4.2 Relative parameters and their values used in the stage 1.

Method	Related Parameters
Alpha range selection	1. *PA*: A percentage of total points with high alpha values 2. The tested value range for *PA*: {0.01, 0.02, 0.04, 0.08, 0.1}
Multifractal analysis	Patch size (for texture enhancement): 1. 128 ×128 pixels (for large microcalcifications) 2. 32 × 32 pixels (for small microcalcificaitons)
Linear structure detector	1. θ: the orientation angle 2. The angle range used for θ: {0°, 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, 330°, 360°}

The proposed MCs detector uses two scales of sliding windows to detect MC spots in different sizes, and detection results are integrated for evaluating the final detection performance (as showed in Fig. 4.2). The INbreast dataset with pixel-level MCs ground truth information is used to test the MCs detector. To train the SVM classifier used in this detector, 150 training patches (32 × 32) are cropped from 10 mammograms, with 50 of them belonging to the MC category and 100 to the normal category.

We use I_{MC1} and I_{MC2} to denote the detection results in 128 × 128 and 32 × 32 sized sub-images, and I_{MC} is the final result combining two parts (i.e. $I_{MC} = I_{MC1} + I_{MC2}$). In order to analyse the experimental results, we define the following rules to count true positive (TP) and FP numbers. In I_{MC1}, if one detected MC spot overlaps the ground truth contour, it is counted as one TP. Otherwise, it is regarded as one FP. In I_{MC2}, since some of the tiny MCs are labelled with only one pixel in mammograms, we form the detection rule that if one detected MC spot is within three pixels from a ground truth point, then this point is counted as one TP. Otherwise, this point is sent to the FP group. We use SN to denote the number of all pixels in the breast region of one mammogram and use GT_n and GT_p to denote ground truth numbers of negative points and positive points, then $GT_n = SN - GT_p$, where GT_p is offered by the INbreast dataset. If TP and FP are observed true positive number and false positive number, then the observed false negative (FN) and true negative (TN) can be computed as follows. Fig. 4.12 illustrates the processing steps in our experiment and one MCs detection result image I_{MC} in a local region.

$$FN = GT_p - TP \tag{4.14}$$

$$TN = GT_n - FP \tag{4.15}$$

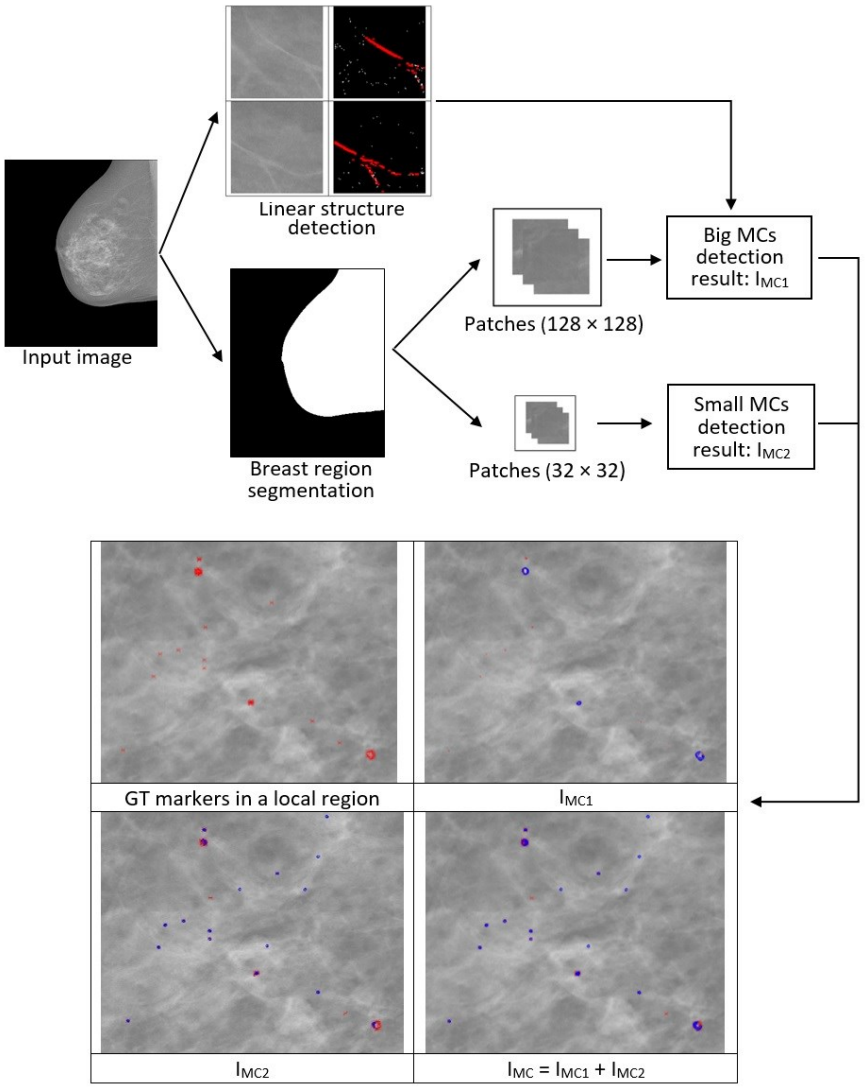


Figure 4.12: Main processing steps in our experiment and an example of MCs detection results in a local region.

In the result image I_{MC}, a threshold value T is set to filter FPs. Here, T means that there should be at least T detected points around a target MC spot. In our experiments, five values (1, 2, 3, 4, and 5) are assigned to T respectively. Confusion matrices with different T values are given in Table 4.3, showing the performance of the proposed method.

Table 4.3 Confusion matrices with different *T* values.

		Predicted		Predicted		Predicted		Predicted		Predicted	
		T=1		T=2		T=3		T=4		T=5	
		Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Actual	Positive	27.2	4.7	25.7	6.2	24.1	7.7	22	9.8	19.8	12.1
	Negative	164	$>10^6$	89.7	$>10^6$	53.1	$>10^6$	32.9	$>10^6$	22.9	$>10^6$
Sensitivity		0.8527		0.8056		0.7579		0.6918		0.6207	
Specificity		0.9998		0.9999		0.9999		1.0000		1.0000	

There are other methods for detecting MCs in the literature (Domingues & Cardoso, 2014) (Zhang, Wang, Li, & Bai, 2014), including Bayesian surprise method, mathematical morphology, and outlier detection, and they are tested using the same dataset. Our methods are analysed and compared with the reported results in (Domingues & Cardoso, 2014). The results comparison in Table 4.4 and the FROC curve analysis in Fig. 4.13 show that our detection results outperform other methods. The Bayesian surprise method shows a better performance than other methods with sensitivity of 60.3% in (Domingues & Cardoso, 2014), but the average FP number (108 per image) indicates that the overall detection accuracy is not satisfactory. By using our method, the sensitivity of 80.6% is achieved, which is much higher than the reported methods and the average FP number (90 per image) is lower. When the average FP number is reduced to 53 by setting $T = 3$, the sensitivity (75.8%) still is the highest among these considered approaches. However, this number of FPs is not acceptable to radiologists' work. Fig. 4.14 gives some examples of detection results using the MCs detector, from which we can see that FPs exist commonly and affect the overall detection accuracy. Therefore, further improvement aimed at FP reduction is proposed and discussed in stage 2 in the proposed framework.

Table 4.4 Results comparison between different schemes reported in (Domingues & Cardoso, 2014) and our method.

Method		Sensitivity (%)	FP (Average number per image)	FN (Average number per image)
Outlier detection		45.8	60	–
Mathematical morphology		40.3	225	–
Bayesian surprise		60.3	108	–
Our proposed method with different threshold values	T=2	80.6	89.7	6.2
	T=3	75.8	53.1	7.7
	T=4	69.2	32.9	9.8
	T=5	62.1	22.9	12.1

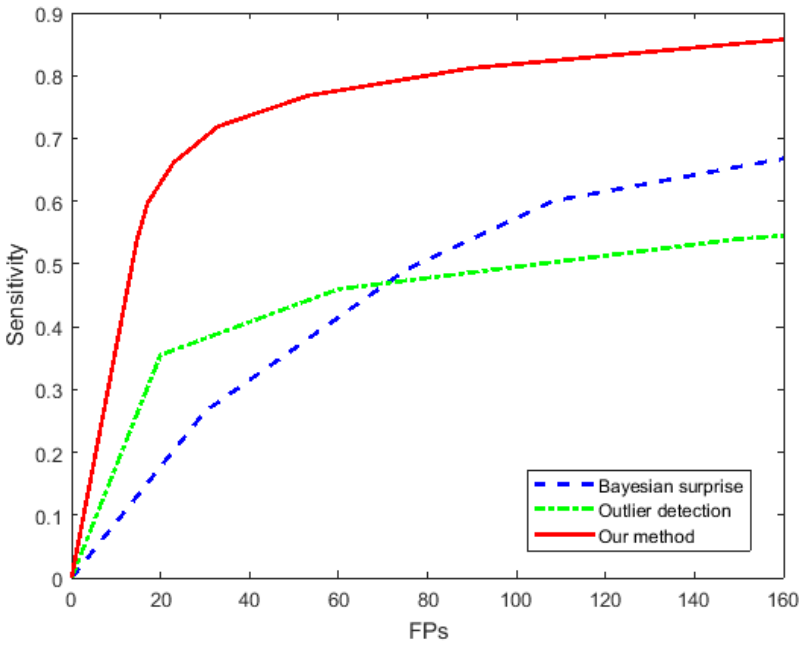


Figure 4.13: FROC curves show the performance of the proposed method.

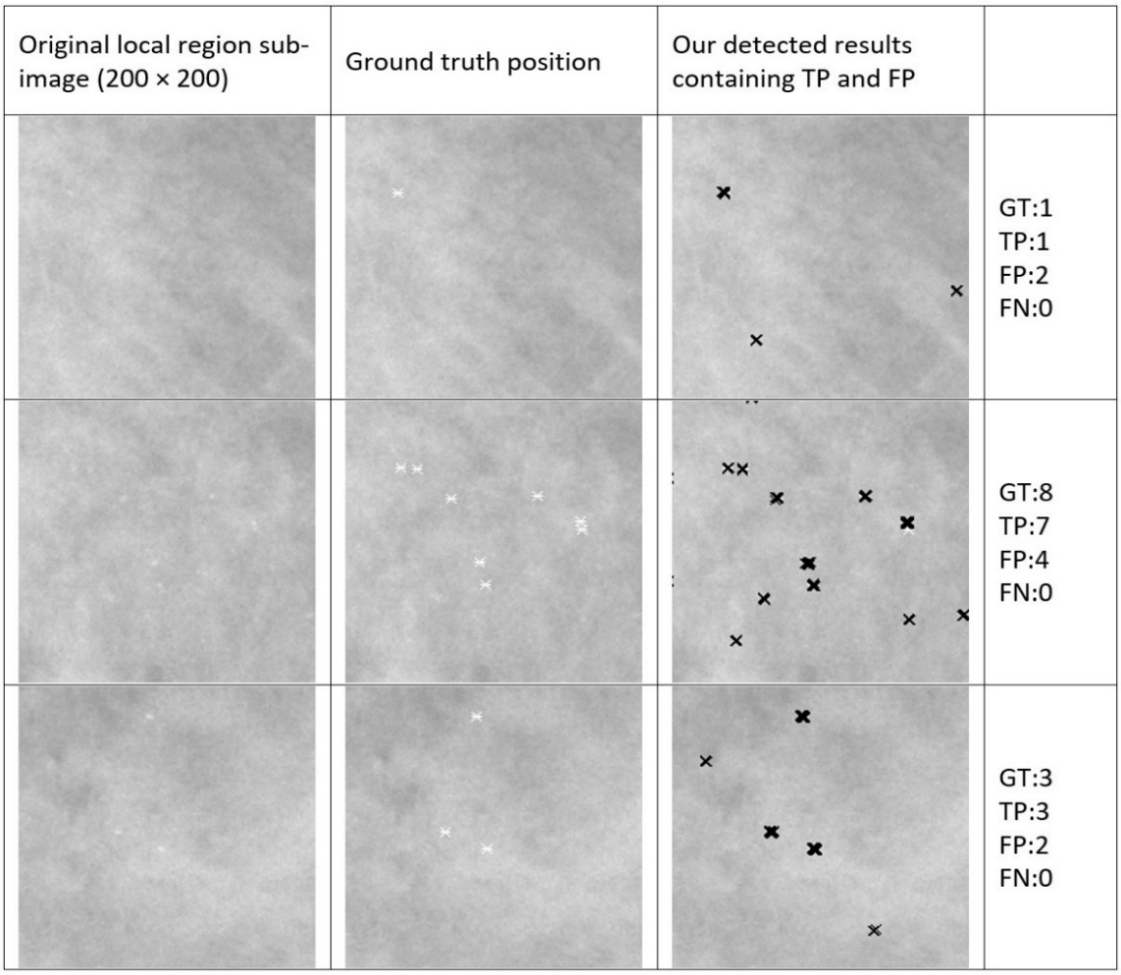


Figure 4.14: Some examples of detection results using the proposed MCs detector, in which the problem of FPs can be seen.

4.4.2 Classification Performance of CNN

Prior to testing the proposed CNN classifier, the MCs detector developed in section 4.2 is used to detect suspicious MC spots from the breast area in mammograms. The detection result contains incorrectly recognized MC spots (FP) and true MCs (TP). Afterwards, small patches with the detected MC spot located in the centre are cropped, constituting a patch set that is fed into the proposed CNN model. The relative parameters and their values used in this CNN classifier (stage 2) are listed in Table 4.5.

Table 4.5 Relative parameters and their values used in the stage 2.

Method	Related Parameters
Multifractal analysis	1. Window edge size: *w* (for calculating alpha values) 2. The tested value range for *w*: {3, 5, 7, 9}
CNN classifier	1. The tested patch sizes: 9 × 9, 17 × 17, 25 × 25, 33 × 33, 49 × 49 (pixels) 2. The convolutional kernel size: 3 × 3 3. The number of filters in each convolutional layer for testing: {16, 20, 32, 40, 64} 4. Batch size: 125 5. Epoch number: 500 6. The learning rate is initialized with 0.001

Due to the imbalance of numbers of TP and FP patches (650 vs. 2733 in the training and validation set), data augmentation is used to enlarge the TP group size: each TP patch is rotated by 90, 180 and 270 degrees, respectively. No augmentation is used on the test set. Table 4.4 shows the detailed distribution used in our experiments.

As shown in Table 4.6, 244 mammogram images (62 cases) containing MCs are divided into training, validation, and test in the ratio 2:1:2, which result in 5333 patches used in training/validation phases and 4111 patches used in test. Note that the input patches are obtained from the MCs detector results (stage 1) rather than original mammograms, so we cannot increase the number of images or patches. However, based on the detected MC position information in stage 1, we cropped small patches with an individual MC spot in the centre of a patch and tested different patch sizes (from 9 × 9 to 49 × 49). Fig. 4.15 shows that the use of 33 × 33 sized patches produces better classification results.

Table 4.6 Distribution of INbreast images and patches used in the proposed method.

	Patch number	TP	FP	Image	Case
Training	3735	1820*	1915	98	25
Validation	1598	780*	818	48	12
Test	4111	452	3659	98	25
Total	9444	3052	6392	244	62

* Augmentation operation is used.

4.4.2.1 Alpha Patches Applied in CNN Classifier

In order to reduce FPs, we propose to use a texture enhanced image (α patch) together with its original patch in the designed CNN model (section 4.3.2). As the

highlighted texture features in the alpha patch depend on the choice of multifractal measures, different local measures and the proposed Weber-based Max measure (WMax) in section 4.3.1 are used separately. In addition, the parameter w, the window size used to calculate α in Eq. (4.1), is analysed in a set of {3, 5, 7, and 9} for obtaining the optimal value that gives the best performance.

The proposed CNN classifier is trained using the stochastic gradient descent (SGD) method with batch size of 125 for 500 epochs. Various unknown parameters are determined using the adaptive moment estimation (Adam) method to minimize the binary entropy loss. The learning rate is initialized to 0.001 and decreased by a factor of 0.5 after every 50 epochs. The FC block in Fig. 4.11 includes two fully connected layers with 600 and 200 neurons, respectively.

In the stage 2, we focus on reducing FP number from the initial detection results of stage 1, and small sized patches (33 × 33) are used in the proposed CNN classifier. Therefore, the depth of the CNN is not as deep as other CNN structures which process larger images (e.g. image size in 300 × 300). Table 4.1 lists 7 different structures tested in this work, with different numbers of convolutional layers considered and different filter numbers applied in each layer. Classification experiment is repeated by using the considered CNN structures, and their classification results are compared in Table 4.7. Experimental results indicate that the fourth structure (#4) in Table 4.1 with two convolutional layers and 40 filters used in each layer produced better classification performance (on TNR and TPR).

Table 4.7 The performance on the test set using different CNN structures.

	Performance (%) on different CNN structures						
	#1	#2	#3	#4	#5	#6	#7
TNR	89.18	87.92	88.72	87.66	85.67	81.97	**95.48**
TPR	78.94	83.07	84.42	**90.00**	88.68	87.38	67.36

In order to evaluate the classification performance of the proposed method, we conduct analysis of true negative rate (TNR), true positive rate (TPR) and accuracy (ACC). High TNR indicates that more FP patches in the input set are classified correctly and overall FPs will be suppressed effectively. Meanwhile, TPR should be kept as high as possible for maintaining a desirable overall detection accuracy. Table 4.8 and Table 4.9 show the results on the training set and the test set, respectively, which indicate that the proposed WMax measure works well, generating the best TPR (w = 9) with slight decreases in TNR and ACC.

Table 4.8 Classification performance on training set.

	w=3			*w*=5		
	TNR	TPR	ACC	TNR	TPR	ACC
Iso	91.06	83.18	87.31	91.29	84.09	87.81
Sum	92.22	84.43	88.44	91.07	92.82	91.94
InvMin	**96.03**	86.51	91.31	92.50	90.77	91.62
Max	89.20	90.32	89.75	93.65	90.02	91.81
WMax	92.60	89.23	90.94	92.06	89.84	90.94
	w=7			*w*=9		
	TNR	TPR	ACC	TNR	TPR	ACC
Iso	95.00	90.39	92.81	91.81	90.03	90.94
Sum	94.21	91.72	93.06	95.55	92.06	**93.88**
InvMin	83.25	92.41	87.63	89.29	91.47	90.38
Max	93.48	90.72	92.12	90.82	92.48	92.63
WMax	92.40	91.96	92.19	92.40	**94.51**	92.44

Table 4.9 Classification performance on test set.

	w=3			*w*=5		
	TNR	TPR	ACC	TNR	TPR	ACC
Iso	74.01	65.66	73.09	84.48	67.67	82.62
Sum	92.49	65.89	89.55	89.66	79.76	88.58
InvMin	**92.87**	70.42	90.43	88.93	73.51	87.28
Max	87.29	81.12	86.61	92.85	75.76	**90.99**
W-Max	90.87	77.37	89.38	90.75	79.41	89.48
	w=7			*w*=9		
	TNR	TPR	ACC	TNR	TPR	ACC
Iso	90.52	73.46	88.61	88.03	74.41	86.56
Sum	91.26	80.05	90.04	92.31	78.67	90.78
InvMin	90.25	75.36	88.66	84.91	74.47	83.77
Max	90.34	81.82	89.40	87.40	83.53	86.97
WMax	89.03	83.06	88.38	87.66	**90.00**	87.92

4.4.2.2 Comparison Using Different Patch Sizes

The size of the input patch should be chosen carefully. If the size is too large, other MCs in the neighbourhood and more background tissue will be included, making texture features complicated. On the other hand, a smaller patch size may not contain sufficient background information to be contrasted with the patch centre. Different sizes (from 9×9 to 49×49 pixels) of the patch are tested in experiments, and a best classification accuracy can be seen when it is set to be 33×33 pixels. Fig. 4.15 shows a comparison of the classification performance using different patch sizes on the validation set. Another experimental find is that this model can reach the best classification accuracy using a training set of 3000 patches (half TP and half FP), and no better result is observed on a larger training set, indicating an overfitting problem.

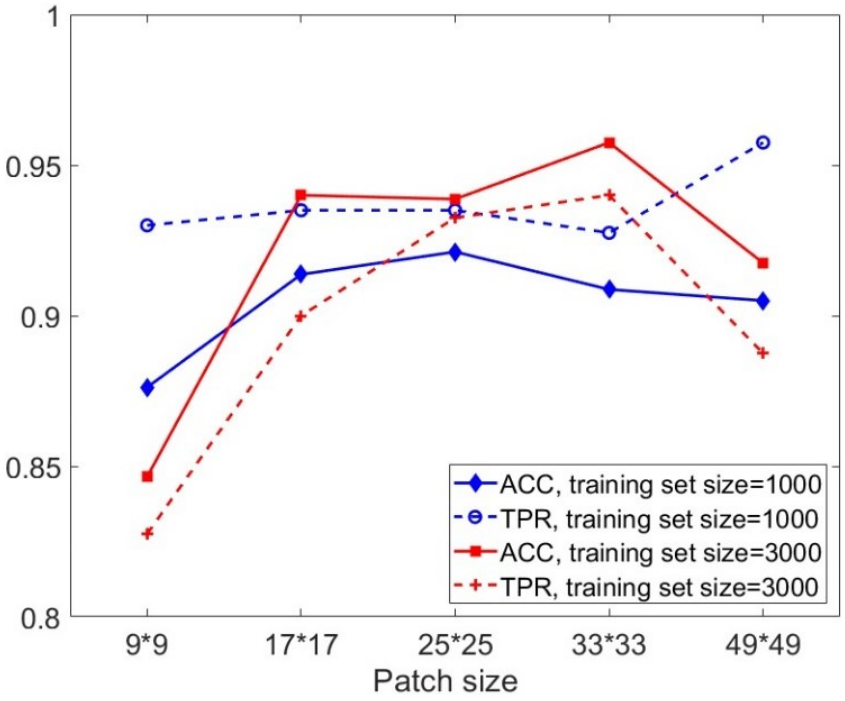


Figure 4.15: The size of the input patch and the training set affect the performance of CNN classifier.

4.4.2.3 Classification Performance Evaluation

In order to demonstrate the effect of multifractal features in alpha patches in this CNN classifier, the original MC patch set and alpha patch set are tested individually. Table 4.10 shows that the classification result is improved after using the two input sets and concatenating their feature maps. Particularly, when the WMax-based alpha patch set is applied, TPR is improved by nearly 15% compared to using Max measure. In addition, another CNN simulating AlexNet architecture is tested. An AlexNet-like network is constructed with 3 convolutional layers, 2 pooling layers and 3 FC layers. In Table 4.8, except the best TNR (92.85%) by using the original MC patch set and WMax-based alpha patch set in the proposed CNN classifier, AlexNet-like network also represent high TNR. However, only high TNR cannot guarantee desirable improvement on overall MCs detection accuracy, which also requires high TPR. We can see that the use of the alpha patch set, individually or together with the original patch set, improves the TPR and without too much decrease in TNR. This demonstrates that multifractal features contained in the alpha images help in recognizing TPs when the model is suppressing FPs.

Table 4.10 Classification performance on test set using different CNN models.

Input set	TNR	TPR	ACC
Alpha patch (Max)	86.45	82.83	86.05
Alpha patch (WMax)	85.67	84.47	85.54
Original MC patch	53.82	44.16	52.74
MC + Alpha (Max)	**92.85**	75.76	**90.99**
MC + Alpha (WMax)	87.66	**90.00**	87.92
MC (AlexNet-like network)	89.16	77.99	87.94

4.4.3 MCs Detection Evaluation

The MCs detector proposed in Section 4.2 and the FP reduction process presented in section 4.3 are integrated to constitute a complete MCs detection framework. To evaluate the overall MCs detection performance, the final detected MC spots are marked in their original mammograms to audit FPs in each mammogram image and FROC is used to analyse the detection results. A FROC curve is a plot of TPR versus the average number of FPs per image with the decision threshold varied over an operating range (Samuelson & Petrick, 2006). Fig. 4.16 gives an example of the final MCs detection results marked in its original mammogram image, and Fig. 4.17 shows a comparison between the initial detection results using the MCs detector and the final results after reducing FPs. The FROC curves in Fig. 4.18 show that the proposed WMax measure combined with the CNN classifier obtained the best MCs detection accuracy, with TPR = 77.9% when the average FP number is under 30 per image. Comparing to the initial results, we can see that the use of the FP reduction process improves the overall detection performance significantly. In addition, the area under the FROC curve (AUCFROC) is computed when the average FPs number is controlled in a range of [0, 30]. The AUCFROC of the WMax-based approach is 0.7034, which is higher than that of the Max-based approach (AUCFROC = 0.6284), indicating that the extended WMax measure works better in this task.

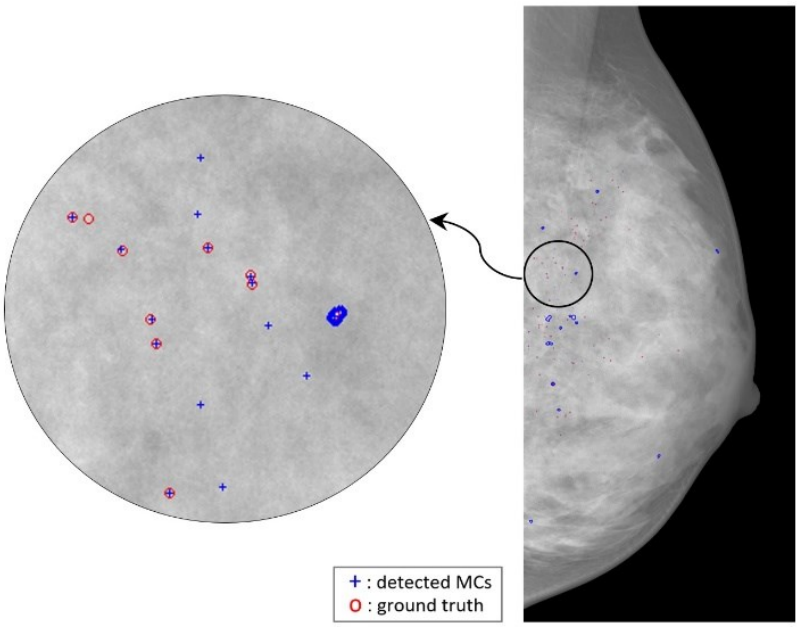


Figure 4.16: The final MCs detection results after replacing the detected individual MC spots in its original mammogram image.

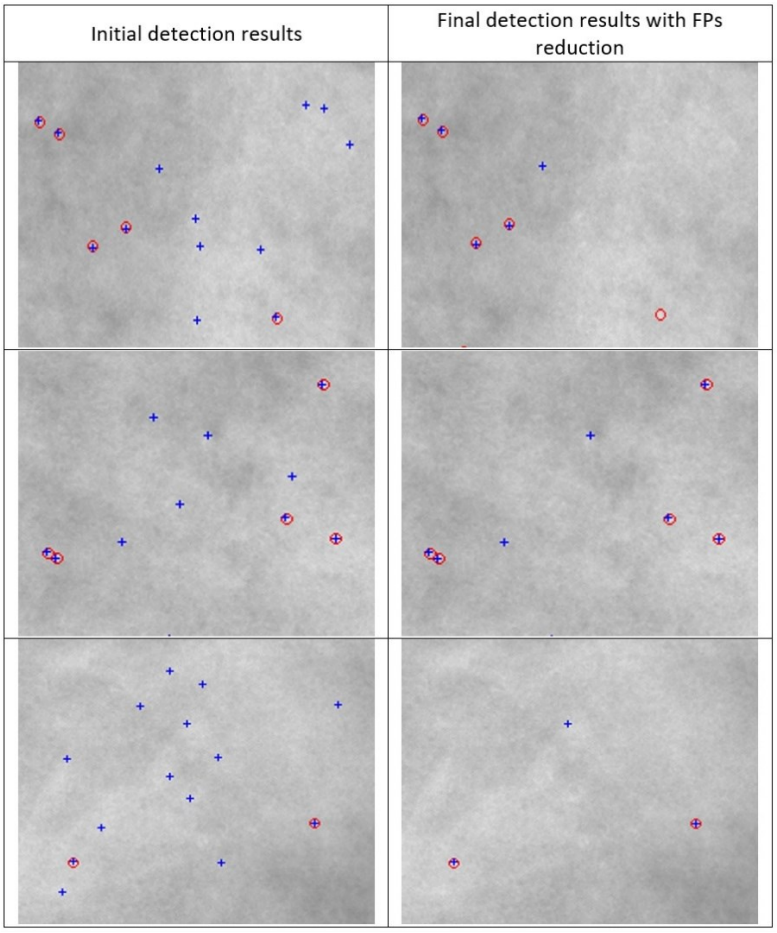


Figure 4.17: Some examples of initial MCs detection results using the proposed MCs detector and the final detection results after combining FP reduction procedures.

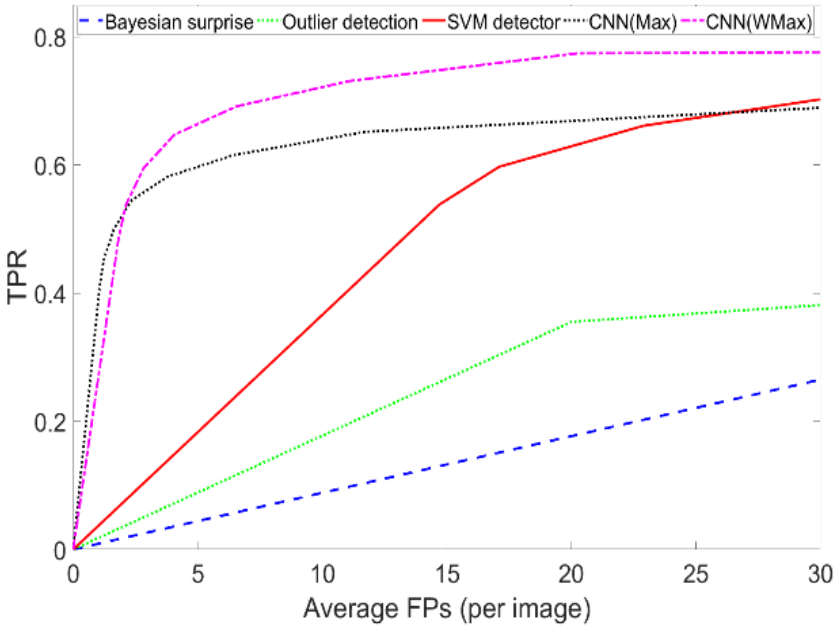


Figure 4.18: FROC curves obtained by different methods.

4.4.4 Robustness of The Proposed Method

As introduced in Section 4.2 and 4.3, the proposed framework for MC detection uses different methods to enhance image textures, filter linear structures and reduce false positive numbers, producing the final detection results. The use of these methods which include multifractal analysis, the linear structure detector and the CNN classifier involves a number of parameters, and the experimental results displayed in Section 4.4 show that the use of different parameters can affect the detection performance differently.

In the proposed method, multifractal analysis is used to enhance the image textures related to MC spots. In addition to considering four commonly used multifractal measures (i.e. Maximum, Inverse-minimum, Summation, and Iso) for calculating the α values, we proposed another local measure called Weber-based Maximum measure (WMax) in Section 4.3.1 with the aim of extracting more informative texture features. Experimental results in Section 4.4.2 show that the sensitivity (i.e. true positive rate) can vary from 74% to 90% based on different measures used in this task, and the highest sensitivity is obtained when using the WMax measure.

Furthermore, we also tested different window sizes when calculating the α values in multifractal analysis. As introduced in Section 4.2.1, a square window of size $w \times w$ is used to estimate the α values. In our experiments, different values (i.e. {3, 5, 7, 9}) of w

are tested, and experimental results in Section 4.4.2 show that the sensitivity increases from 77% to 90% when different w values are set.

In the proposed method, a small patch size is used to detect tiny MC spots (diameter less than 5 pixels). In our experiments, the patch size is set in a range of {9 × 9, 17 × 17, 25 × 25, 33 × 33, 49 × 49 pixels} to find out the optimal size in this detection work. Experimental results in Fig. 4.15 show that the sensitivity reaches the highest point (approximately 93% on the training set) when the patch size is 33 × 33, comparing to the sensitivity of 83% when using the 9 × 9 sized patches.

In the proposed CNN classifier, different network structures (Table 4.1) are tested. In each convolutional layer, the number of filters is set in a range of {16, 20, 32, 40, and 64}. The experimental results in Table 4.7 shows that the CNN structure using 40 filters in each convolutional layer produces the best sensitivity (90%), while the lowest sensitivity is under 70%. The use of more filters (e.g. 64 filters in structure 5) or a much deeper structure (e.g. structure 7 in Table 4.1) does not give a better detection performance.

Through the discussion above, we can see that although the proposed MC detection framework produces desirable MC detection results comparing to other methods, it involves a number of parameters and a problem of how to choose their values properly. For example, the use of different multifractal measures or different window size (w) can cause approximately 15% difference on the sensitivity. Particularly, the current experiment is based on a specific mammogram dataset (i.e. INbreast) which is the only publicly available dataset with the pixel-level ground truth of MCs. We can presume that other mammogram images produced using different devices possibly have different image resolutions, which will require repeating the experiment to decide the optimal parameters. Further work will be planned to test the robustness of the proposed framework after obtaining other mammogram datasets.

4.5 Chapter Summary

This chapter introduces a complete MCs detection framework containing two stages: MCs detection and a FPs reduction process, with the aim of detecting MC spots in mammograms. The proposed MCs detector integrates a multifractal texture enhancing process in two scales of sub-images for detecting MCs of different sizes. A linear structure filter is developed to distinguish linear arrangement of tissue from MCs and a SVM classifier is trained using the feature vector containing the selected α-values and

pixel intensity information to generate the detection results. To reduce the number of false positive and improve the overall detection accuracy, this chapter also addresses a novel Weber's law based multifractal measure and a patch-wise CNN classifier. The proposed CNN classifier processes both original mammogram patches and corresponding alpha patches which feed multifractal features into the model and help in learning more effective image features related to MCs. A digital mammogram dataset, INbreast, is used to test the proposed framework, and detection results are analysed by free-response receiver operating characteristic (FROC). Experimental analysis demonstrates that the proposed detection framework outperforms other MCs detectors with a higher sensitivity and lower FP number.

5 BREAST DENSITY EVALUATION USING MULTIPLE IMAGE FEATURES

This chapter introduces two breast density classification models which focus on the use of multiple image features for improving classification performance. The considered image features in the two models include multifractal spectrum, local binary patterns, and texture feature based histograms. In addition, the first model uses a chi-square test statistic to classify cropped mammogram patches to dense and fatty categories, based on which percentage density is calculated for classification. The second classification model adopts an autoencoder network and principal components analysis to optimise cascaded image features, classifying mammograms into different density categories. INbreast and MIAS datasets are used to test the proposed classification models.

5.1 Breast Density and Its Classification Criteria

Breast density indicates the amount of fibrous and glandular tissue compared with the amount of fatty tissue in a woman's breasts. Mammographic density (MD) classification relates to measuring the amount of fibro-glandular tissue in mammogram images and classifying them into different categories. Different measurement metrics have been proposed and used to evaluate breast density, including Six-Class-Categories (SCC) (Boyd et al., 1995), Wolfe's four categories (Wolfe, 1976) and Breast Imaging-Reporting and Data System (BI-RADS) (Sickles et al., 2013). The BI-RADS breast

density criterion proposed by the American College of Radiology (ACR) has been widely used in clinical applications and includes four density categories: fatty, scattered density, heterogeneously dense, and extremely dense. BI-RADS (4th edition) also gives the correspondence between the density category and the proportion of fibro-glandular tissue area within the breast region in mammograms (Table 5.1). Women in the first two categories are said to have low-density or fatty breasts. Women in the second two categories are said to have high-density or dense breasts. Women with dense breasts have a higher chance of getting breast cancer (Mohamed et al., 2018).

Table 5.1 BI-RADS breast density classification criterion.

BI-RADS Category	Percent Density
BI-RADS I : almost entirely fat	< 25% glandular density
BI-RADS II : scattered fibro-glandular densities	25—50% glandular density
BI-RADS III : heterogeneously dense	51—75% glandular density
BI-RADS IV : extremely dense	> 75% glandular density

5.2 Classification Model Using Multifractal Spectrum

This section introduces a density classification model which uses multifractal spectrum to classify mammogram patches into dense or fatty categories and calculate the percentage density. Multifractal spectrum contains a series of fractal dimension values corresponding to sets of points with a singularity exponent (α value), and can be used to describe texture features in an image. This model calculates multifractal spectrum for each cropped local region in mammograms to reflect texture features related to dense fibro-glandular tissue or fatty tissue. Histogram analysis based on the original breast area is conducted to obtain a proper α value range for producing effective multifractal spectra in this classification work. Fig. 5.1 shows a pipeline of the proposed model, and more details are given in the following sections.

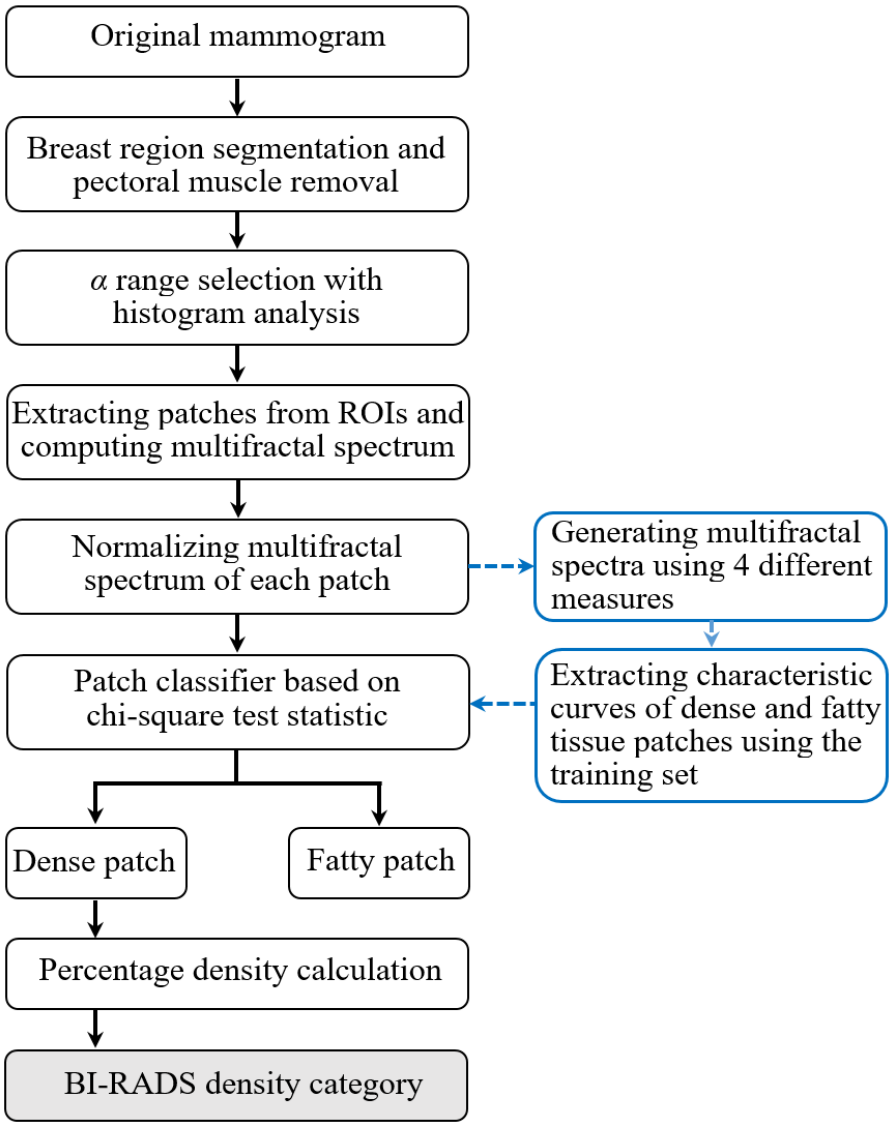


Figure 5.1: The pipeline of the proposed classification model.

5.2.1 Multifractal Spectrum

Multifractal spectrum can be used as a global feature descriptor for image classification and interpretation, and prior research work has demonstrated its effectiveness in medical image analysis (Ibrahim & Mukundan, 2014)(Paskas, Reljin, & Reljin, 2015) (Paskas, Reljin, & Reljin, 2016). However, this technique has not been used to process and analyse mammogram images for breast density classification. Based on the introduction of multifractal analysis in Section 4.2, the subdivision of α-value range of an input image gives a decomposition of the image in terms of a set of α-slices. By calculating fractal dimension ($f(\alpha)$) of each α-slice, a plot of multifractal spectrum can be obtained with α-slices along the x-axis and $f(\alpha)$ along the y-axis. These plots through discrete points are often interpolated to form a continuous curve for easier

reading and analysis. A box-counting method (Bouda, Caplan, & Saiers, 2016) is used in this work for calculating the fractal dimension due to its simplicity. The box-counting method counts the number of boxes containing at least one pixel with its α value belonging to a specific α-interval range $[\alpha_i, \alpha_{i+1}]$. α intervals are obtained by dividing the α range $[\alpha_{min}, \alpha_{max}]$ into a pre-specified number (N) of sub-ranges. Let $n(\varepsilon)$ denote the number of boxes containing at least one pixel with its α value belonging to a specified α-interval, where ε is the box size, then the fractal dimension is given by:

$$f(\alpha) = \lim_{\varepsilon \to 0} \left(\frac{\log(n(\varepsilon))}{\log(\varepsilon)} \right) \tag{5.1}$$

Fig. 5.2 illustrates four α images and their multifractal spectra using different multifractal measures.

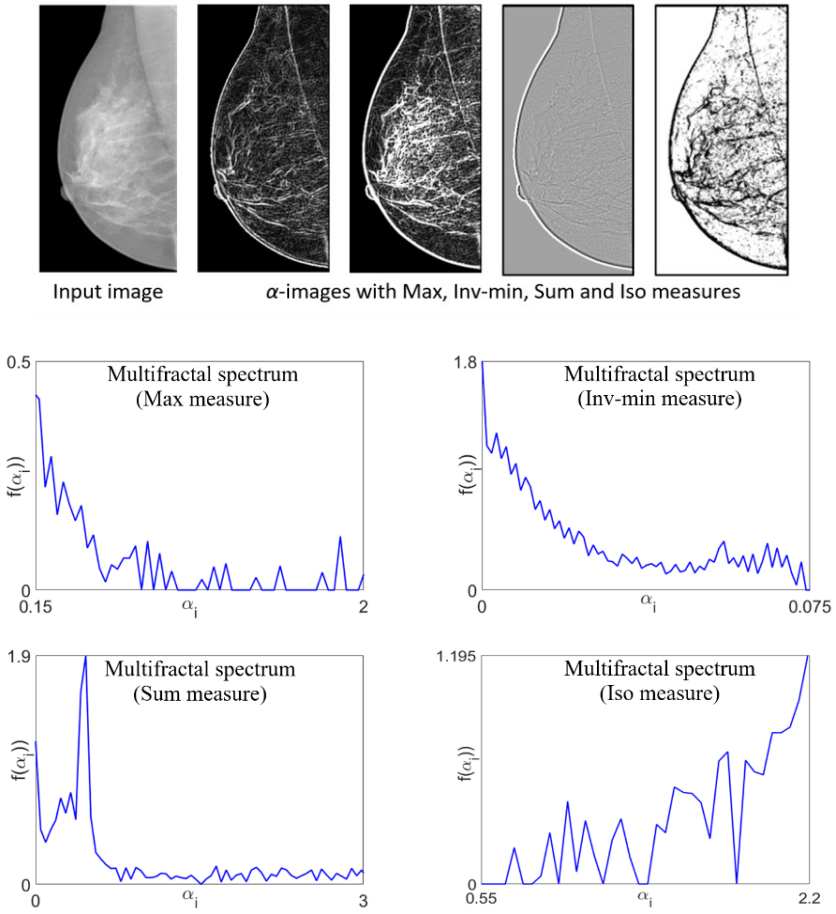


Figure 5.2: Examples of α images and their multifractal spectra.

5.2.2 Alpha Range Selection Based on Histogram Analysis

As discussed in section 4.2, using a specific α value range instead of its entire range produces a better texture enhancement result. Therefore, we need to develop appropriate α-value selection rules, based on which effective multifractal spectra can be obtained.

For each mammogram, its α image and the whole α value range [α_{min}, α_{max}] are generated first, then a specific percentage (PA) of total points with lower α values in the ROI are retained for deciding a proper α value range as follows.

$$\sum_{i=\min}^{t-1} n(\alpha_i) = PA \times Total\ pixel\ number\ in\ ROI,\ t \in [\min, \max] \tag{5.2}$$

where, $n(\alpha_i)$ denotes the number of points possessing the α value in the range of [α_i, α_{i+1}]. The α image in range of [α_{min}, α_t] is expected to show better texture patterns related to fibro-glandular tissue areas. Fig. 5.3 gives an example to show that the use of different PA values decides corresponding α value ranges and also affects the texture enhancement process (Fig. 5.3). Therefore, we do not set a unified threshold value for PA, instead an adaptive method is developed subsequently for obtaining useful α images and multifractal spectra.

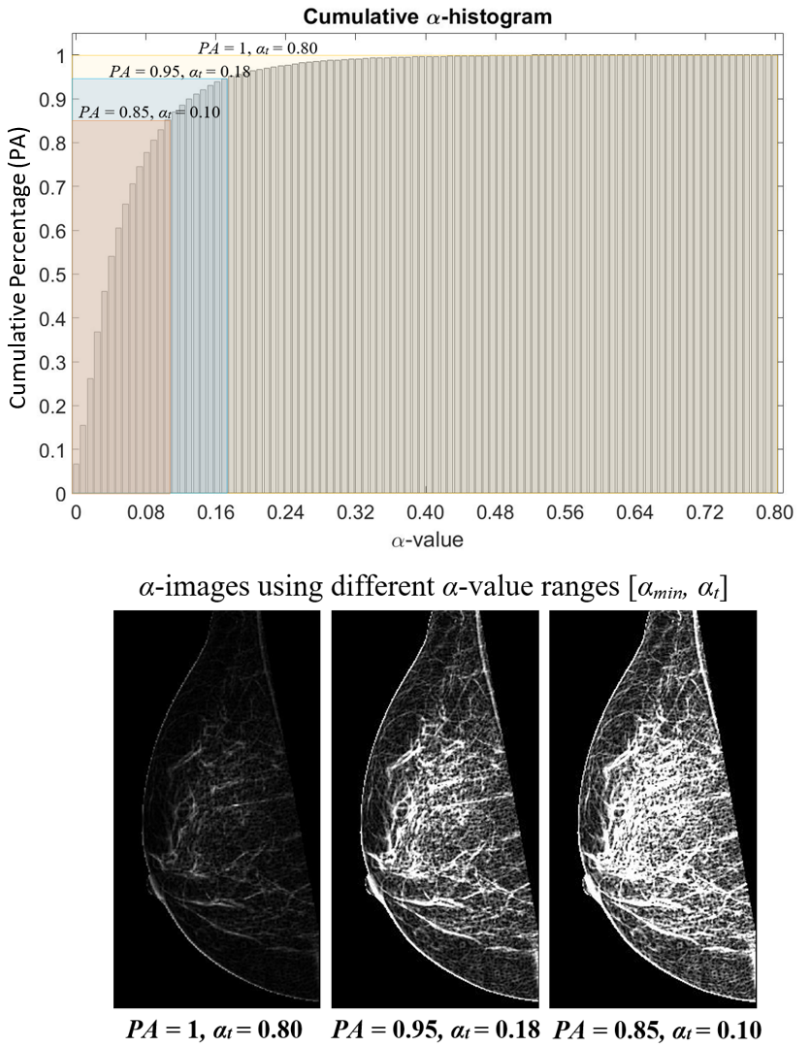


Figure 5.3: A cumulative α-histogram of the mammogram image in Fig. 5.2 and its α-images (Inv-min measure) using different α-value ranges decided by PA.

In this work, histograms based on intensity information in breast areas are analysed, as the shapes of the histograms display some characteristics related to different breast densities (Fig. 5.4). Concretely, histograms of entirely fatty breasts (i.e. BI-RADS I) present higher peaks than the histograms of dense breasts, while histograms of extremely dense breasts (i.e. BI-RADS IV) have lower peaks and show skewed distribution to the left compared to breasts with scattered or heterogeneous dense areas (i.e. BI-RADS II and III). Note that the histograms discussed here (as shown in Fig. 5.4) are normalized by two operations: i) the height values of bars in a histogram are divided by the breast region area, which means the height of each bar in the histogram does not denote the number of pixels but represents the percentage that corresponding pixels account for in the breast region; ii) the bars covering the pixels with the lowest 5% intensity values and the highest 5% intensity values are cut off, as they always contain outliers. Standard deviation and skewness of the histograms are used to design adaptive α value selection rules. Standard deviation (*SD*) and skewness (*SK*) are calculated as follows.

$$SD = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N-1}} \tag{5.3}$$

$$SK = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^3}{\sqrt{[\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2]^3}} \tag{5.4}$$

where x_i, $i = 1, 2, \ldots, N$, is the percentage of pixels with intensity values in the ith bin; μ is the mean of all observed percentages. In this work, *SD* is used to select the α-value range for entirely fatty breast mammograms and a parameter s in Eq. (5.5) which is empirically determined by combining *SD* and *SK* helps to develop another α-value selection rule for extremely dense mammograms.

$$s = SD \times 10 + SK \tag{5.5}$$

Threshold values for *SD* and s are decided from the possible value range of [0.05, 0.15] and [0, 0.8], respectively, after testing more mammograms in a training set. Fig. 5.5 shows the difference of the values of standard deviation and the parameter s between mammograms in different density categories in the training set. By combining the parameters (*SD*, *SK,* and s) discussed above, a more flexible percentage parameter *PA* in Eq. (5.2) is obtained as follows.

$$PA = \begin{cases} 0.95, & if\ SD \geq 0.09 \\ 0.85, & if\ SD < 0.09, s \geq 0.4 \\ 0.75, & if\ SD < 0.09, s < 0.4 \end{cases} \quad (5.6)$$

The adaptive α-value selection rules with histogram analysis are used to generate α images with enhanced texture patterns and to produce corresponding multifractal spectra for classification.

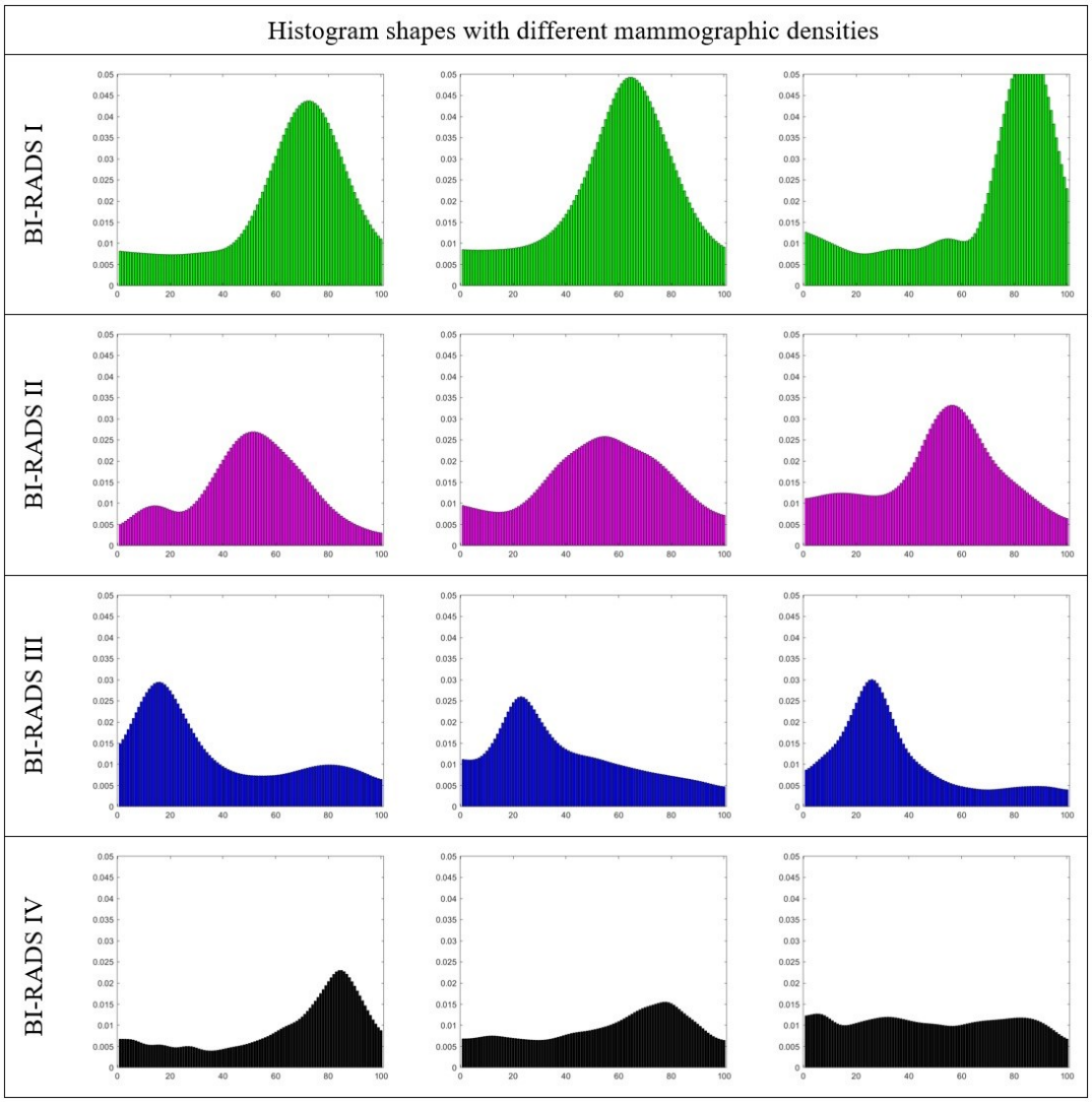


Figure 5.4: Typical histogram patterns of breast areas with different density categories.

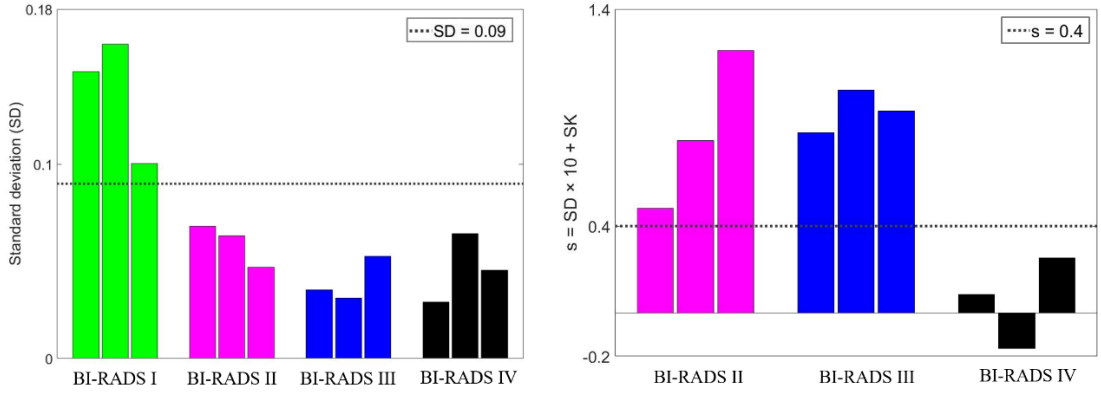


Figure 5.5: The comparison of values of standard deviation and parameter s between mammograms in different density categories.

5.2.3 Characteristic Curves of Breast Density

As previously shown, using α images, mammographic texture features are enhanced and fibro-glandular tissue can be observed clearly. In order to evaluate breast density category in accordance with BI-RADS criteria, dense fibro-glandular tissue regions are recognized first from the background breast region, which can be used to calculate percentage density for classification. Although multifractal spectrum characterises global image features effectively, a spectrum based on the whole mammogram image containing the background area, pectoral muscle region, and various other tissue components cannot characterise fibro-glandular tissue features precisely. Therefore, a patch-wise multifractal spectrum analysis is designed in our proposed method with the aim of distinguishing dense fibro-glandular tissue from fatty tissue. In corresponding α images, after segmenting the breast region and removing the pectoral muscle region, a series of patches of size of 48 × 48 pixels are extracted from breast regions with 50% areas overlapping with each other. All the patches are manually allocated into two groups by an experienced radiologist, dense or fatty patch group, which constitute a training set with 400 patches in each group using 20 mammograms. For each patch, its multifractal spectrum is calculated with the procedure introduced in section 5.2.1 and each spectrum is normalized into an α-value range ([0, 1]). In order to compare different spectra between the two groups, two average multifractal spectrum curves over the training set are displayed in one graph. Such an operation is repeated 4 times using four different multifractal measures (Fig. 5.6).

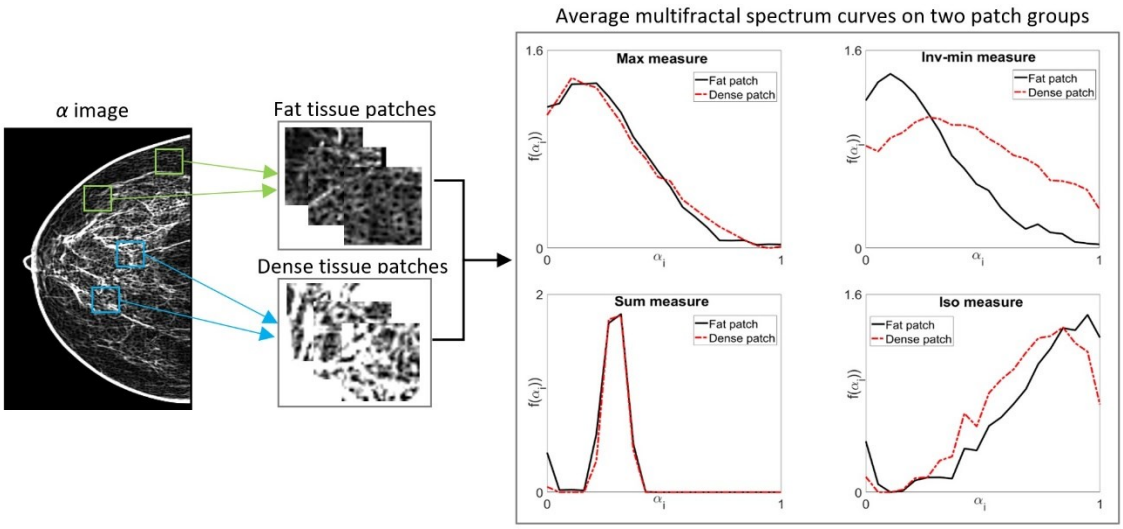


Figure 5.6: Characteristic curves extracted from average multifractal spectra.

In Fig. 5.6, Inv-min measure shows more distinct spectrum patterns than the other measures for characterising the two patch groups. Therefore, Inv-min measure is used to generate the α image with the aim of enhancing fibro-glandular tissue areas and classifying breast density in this model.

5.2.4 Chi-square Test Statistic

A chi-square statistic is used to show the relationship between two categorical variables, and a low value of chi-square indicates that there is a high correlation between the observed and the expected values. In this proposed model, the feature vector of each patch contains 20 elements (i.e. $f(\alpha_i)$ values) which are extracted along the x axis (α_i) of its multifractal spectrum. The Chi-square statistic (χ^2) is calculated as follows.

$$\chi^2 = \frac{1}{2}\sum_{i=1}^{20}\frac{(f(\alpha_i) - f'(\alpha_i))^2}{f(\alpha_i) + f'(\alpha_i)} \tag{5.7}$$

where, each $f(\alpha_i)$ is extracted from the normalized multifractal spectrum of the current patch with equally spaced α_i positions and $f'(\alpha_i)$ is the corresponding fractal dimension value sampled from the Inv-min characteristic curves (Fig. 5.6). Let χ_D^2 and χ_F^2 represent the chi-square statistic results using dense and fat characteristic curves as the expected values, respectively; then if χ_D^2 is lower than χ_F^2, the current patch is classified to dense fibro-glandular category.

5.2.5 Experiments and Results

This section discusses experimental results by using the classification model based on the patch-wise multifractal spectrum technique introduced in Section 5.2. In this model, the histogram analysis based α-value range selection rules Eqs. (5.2) − (5.6) are used to generate α images, and a sliding window (48 × 48 pixels) is used to divide the breast region into a number of patches. For each patch, multifractal spectrum is computed and a feature vector is extracted and input into a chi-square statistic based classifier for classifying the patches into dense or fatty groups. Those classified dense patches together constitute a dense tissue area, which can be used to compute the breast percentage density (PD) for each mammogram as follows. After calculating percentage density, each mammogram could be assigned a breast density label in accordance with the BI-RADS criteria. Fig. 5.7 gives an overview of the processing steps in this model. Relative parameters used in this model are summarised in Table 5.2.

$$PD = \frac{Dense\,tissue\,area}{Breast\,region\,area} \tag{5.8}$$

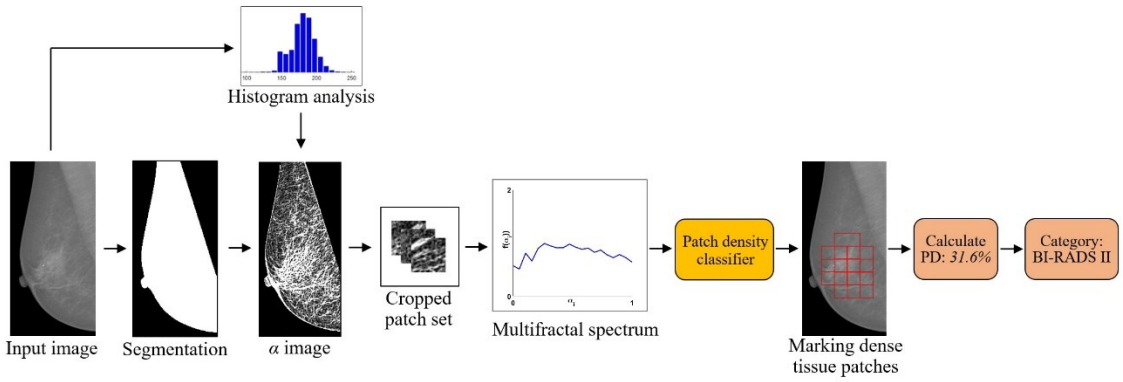


Figure 5.7: An illustration of the main processing steps in the proposed classification model.

Table 5.2 Important parameters used in the proposed classification model.

Method	Related Parameters
Alpha range selection	1. A percentage (*PA*) of total points with lower alpha values is adaptively decided. 2. The tested range for the threshold value of standard deviation: [0.05, 0.15] 3. *s*: a parameter for deciding the value of *PA* (Eq. 5.5)
Chi-square test statistic	1. χ^2_D : the chi-square test value by using dense tissue patches 2. χ^2_F : the chi-square test value by using fat tissue patches
Breast density classification	The threshold values used (based on the BI-RADS criterion): 1. BI-RADS I: $PD \leq 0.25$ 2. BI-RADS II: $0.25 < PD \leq 0.5$ 3. BI-RADS III: $0.5 < PD \leq 0.75$ 4. BI-RADS IV: $0.75 < PD$

In our experiments, 20 mammograms in the INbreast dataset are used as the training set and 78 mammograms corresponding to 18 cases (there are 3 cases that have only one side (left or right) breast views) are used as the test set. Fig. 5.8 shows sample outputs of the key processing stages for each of the density categories. Table 5.3 gives a confusion matrix of the classification results on the test set using the proposed model, in which a classification accuracy of 83.33% is obtained. For 21 cases (patients) in the test set, after adopting a majority vote policy and considering the average percentage density of different views, 19 patients' breast density have been classified into correct groups, with accuracy of 90.1% (19/21).

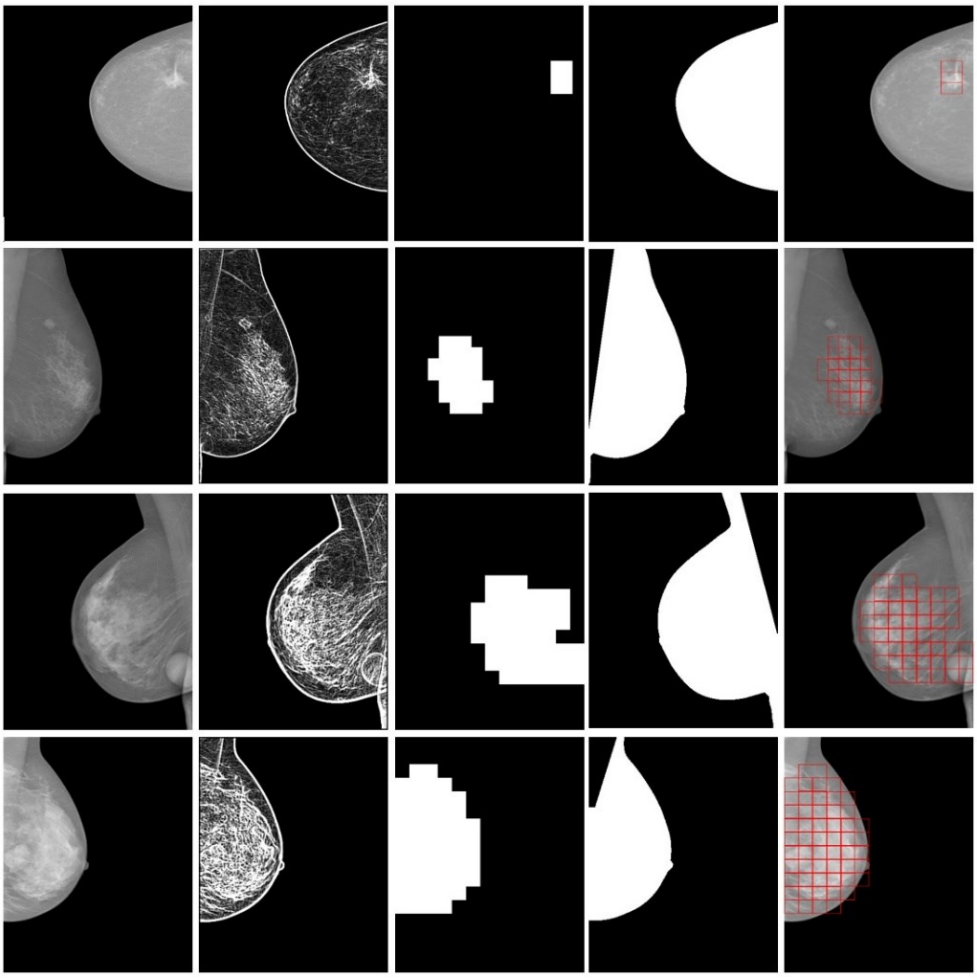


Figure 5.8: Processing results with mammograms in four different density categories. First column: original mammograms; second column: α images using adaptive α value range selection; third column: the combination of classified dense patches using multifractal spectrum; fourth column: segmentation of breast region; fifth column: dense tissue areas are marked in original mammograms.

Table 5.3 Confusion matrix of classification results using the proposed model based on multifractal spectrum.

		Predicted				Accuracy=83.33%	
		ACR-1	ACR-2	ACR-3	ACR-4	**Precision**	**Recall**
Actual	ACR-1	14	2	0	0	0.70	0.88
	ACR-2	6	10	1	1	0.83	0.56
	ACR-3	0	0	22	2	0.92	0.92
	ACR-4	0	0	1	19	0.86	0.95

5.3 Classification Model Using LBP and Autoencoder Network

This section presents a novel breast density classification model using cascaded image features. In order to explore more important and effective texture features, local binary patterns (LBP) and α-histograms based on multifractal measures (introduced in Section 4.2.2) are concatenated to form a new feature vector which is used to classify mammographic density. This model uses principal component analysis (PCA) and an autoencoder network to optimise the feature vector and reduce feature dimensionality. An SVM classifier is trained and used to predict breast density labels of mammograms in the INbreast dataset. Experimental results demonstrate that this model uses the cascaded feature vector to capture more local texture information and presents a competitive classification accuracy. Fig. 5.9 gives a processing pipeline containing the main steps in the proposed classification model.

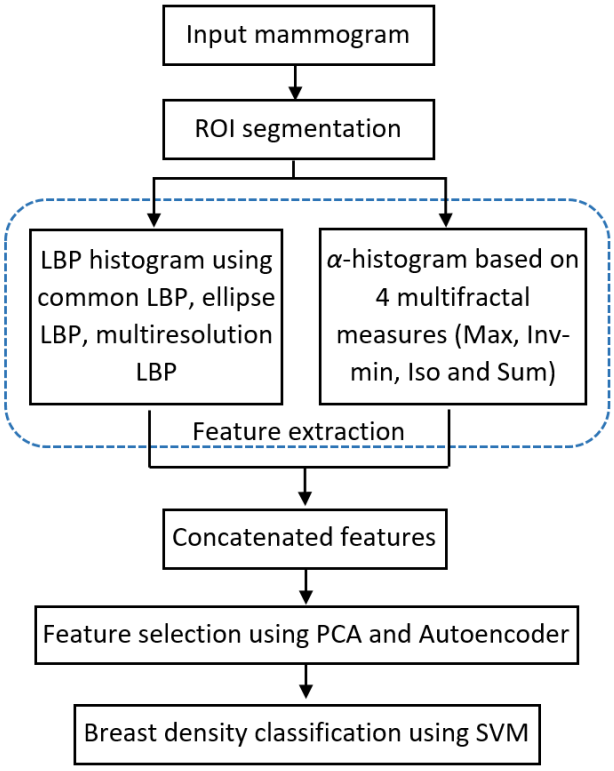


Figure 5.9: Processing steps of the proposed classification model using concatenated features.

5.3.1 Local Binary Patterns

Local binary pattern (LBP) proposed by (Ojala, Pietikäinen, & Harwood, 1996) is a powerful feature descriptor used for texture analysis and classification. Due to its

simplicity and robustness, several research works and applications use it to extract image features. For mammographic density classification, LBP and its variants have been applied and tested to improve the classification accuracy in (Mohamed et al., 2018) (Rampun et al., 2018). A binary pattern is derived by comparing the intensity at each pixel with its neighbours and encoding the information in a P-bit integer value. Concretely, for each central pixel c with a grey level value g_c in a specific window size, its LBP code is calculated by comparing the g_c value with its neighbourhood pixels which is located at a distance R from c. If g_c is higher than the neighbouring pixel P_i, then the neighbour pixel will be assigned a value 0, otherwise value 1 would be allocated. Subsequently, a P-bit binary code is generated for the current pixel c. $LBP^{P,R}$ is used to denote this binary code and its calculation can be described as follows:

$$LBP(c,R,P) = \sum_{i=0}^{P-1} s(g_i^{P,R} - g_c)2^i \tag{5.9}$$

$$s(x) = \begin{cases} 1, & if\ x \geq 0 \\ 0, & if\ x < 0 \end{cases} \tag{5.10}$$

When P is set to 8, 256 (2^8) different binary patterns can be generated using LBP; therefore, an LBP histogram containing 256 bins is obtained and used as a local texture descriptor in this classification model. In addition, for including more local image information, higher R values are used to consider a larger neighbourhood area and generate a longer texture vector after cascading their histograms together. For example, when P = 8, setting R = 1 and R = 2 separately, two LBP histograms are concatenated, producing a 512-length feature vector. This process is called multi-resolution LBP (MLBP), which contains more texture information but increases the feature dimensionality significantly.

In basic LBP, a circular neighbourhood with a radius of R is used for locating neighbouring pixels. As a variant of LBP, elliptical LBP (ELBP) is developed in (Mohamed et al., 2018), aiming to extract more local information from different directions. In ELBP, R_1 and R_2 denote the lengths of the semi-minor and semi-major axis of the ellipse, and ELBP can be calculated as follows:

$$ELBP(c,P,R_1,R_2) = \sum_{i=0}^{P-1} s(g_i^{P,R_1,R_2} - g_c)2^i \tag{5.11}$$

Fig. 5.10 illustrates the basic LBP and its variants discussed above, and they will be used and compared in this classification model.

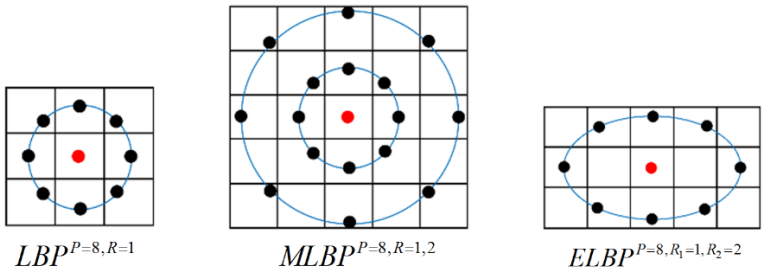


Figure 5.10: Illustrations of basic LBP, MLBP and ELBP.

5.3.2 Autoencoder Network

Autoencoder network (Kramer & Mark, 1991) is a feed-forward neural network with more than one hidden layer, attempting to reconstruct input data at the output layer. The output layer is usually of the same size as the input layer and the network architecture represents an hour-glass shape. As the size of the hidden layer in an autoencoder neural network is smaller than the input layer, the high-dimensional input data can be reduced to narrower code space when using more hidden layers. Therefore, in addition to image reconstruction and compression (Bai, Dai, Wu, & Xie, 2018), autoencoder is also used to reduce feature dimensionality. Generally, an autoencoder network consists of two components, namely "encoder" and "decoder". By reducing the hidden layer size, the encoder part is forced to learn important features of the input data, and the decoder part reconstructs the original data from the generated feature code. Once the training phase is over, the decoder part is discarded and the encoder is used to transform a data sample to a feature subspace.

In this model, we use an autoencoder to reduce the length of the feature vector which concatenates LBP histograms and α-histograms. Therefore, a feature vector of size ($1 \times n$) is input into the autoencoder network for reducing its dimensionality (n). An autoencoder model containing 11 hidden layers is designed as illustrated in Fig. 5.11, in which the input layer receives and processes the initial texture features. Fully connected (FC) layers are used as hidden layers in this proposed structure and rectified linear unit (ReLU) is used as the activation function in each hidden layer except the last layer which uses a sigmoid function. Binary cross entropy is employed as a loss function.

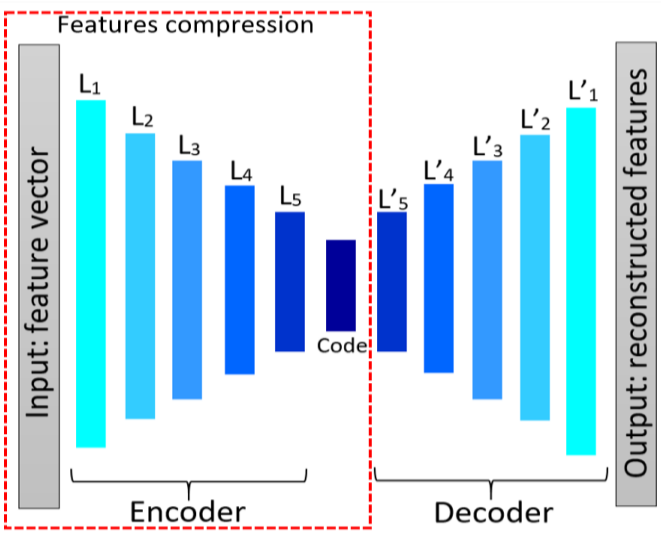


Figure 5.11: The autoencoder network architecture used for reducing the feature dimensionality.

5.3.3 Principal Components Analysis

Principal components analysis (PCA) is used for efficient coding of various biological signals (Hyvarinen, Karhunen, & Oja, 2001). It is a well-known optimal linear scheme for dimension reduction in data analysis, which retains maximal variance in the data set, while improving algorithm performance and saving processing time.

In the proposed method, X is used to denote the input feature set which cascades histogram information. X is an $M \times N$ matrix which has N dimensional features and M elements in each dimensionality. We use X^i to refer to the entire set of elements in the ith dimension and X_j^i to refer to the jth element in this set. Then, the covariance between two dimensions can be calculated as below, which computes the covariance between the first two dimensions X^1 and X^2.

$$\mathrm{cov}(X^1, X^2) = \frac{\sum_{j=1}^{M} (X_j^1 - \overline{X^1})(X_j^2 - \overline{X^2})}{M - 1} \tag{5.12}$$

where, $\overline{X^1}$ and $\overline{X^2}$ denote means of the set of X^1 and X^2. After computing all the possible covariance values between different dimensions, a covariance matrix CM can be obtained like this:

$$CM = \begin{pmatrix} \mathrm{cov}(X^1, X^1) & \cdots & \mathrm{cov}(X^1, X^N) \\ \vdots & \ddots & \vdots \\ \mathrm{cov}(X^N, X^1) & \cdots & \mathrm{cov}(X^N, X^N) \end{pmatrix} \quad (5.13)$$

Eigenvalues and eigenvectors of the covariance matrix are calculated subsequently, and eigenvectors are sorted in descending order according to the eigenvalues. A matrix V can be constructed with these eigenvectors in columns. The final feature set X' can be derived from X and the matrix V as follows:

$$X' = V^T \times X^T \quad (5.14)$$

In PCA, an assumption made for feature selection is that most information of the input feature set is contained in the subspace spanned by the first n principal axes, where $n < N$ in an N-dimensional feature space. Therefore, each original feature vector can be represented by its principal component vector with the dimensionality of n.

5.3.4 Classification

SVM is used as the classifier to predict the breast density labels for test mammograms. One key point is to find a proper kernel and the optimal parameters (gamma and C). A radial basis function (RBF) kernel is selected as it is used and recommended in (Fan, Chen, & Lin, 2005) (Gangeh et al., 2010) as the first choice. The parameters of gamma (kernel width) and C are found using a grid-search procedure.

5.3.5 Experiments and Results

In order to describe texture features related to breast density, only pixels in the ROI (i.e. the breast region) are considered when generating LBP histograms and α-histograms. The area of the ROI varies in different cases, which indicates that counting the number of pixels for plotting histogram bins cannot reflect texture differences accurately among mammograms. Therefore, a percentage value that the counted pixels account for in the ROI is used to generate the corresponding histograms based on LBP and α values. Fig. 5.12 shows examples of the LBP histogram and α-histograms.

The INBreast dataset containing 409 mammograms with 4 BI-RADS density categories is used to test the proposed classification model. We use a ten-run ten-fold

cross validation method to produce classification results. Relative parameters and their values tested in the proposed model are listed in Table 5.4.

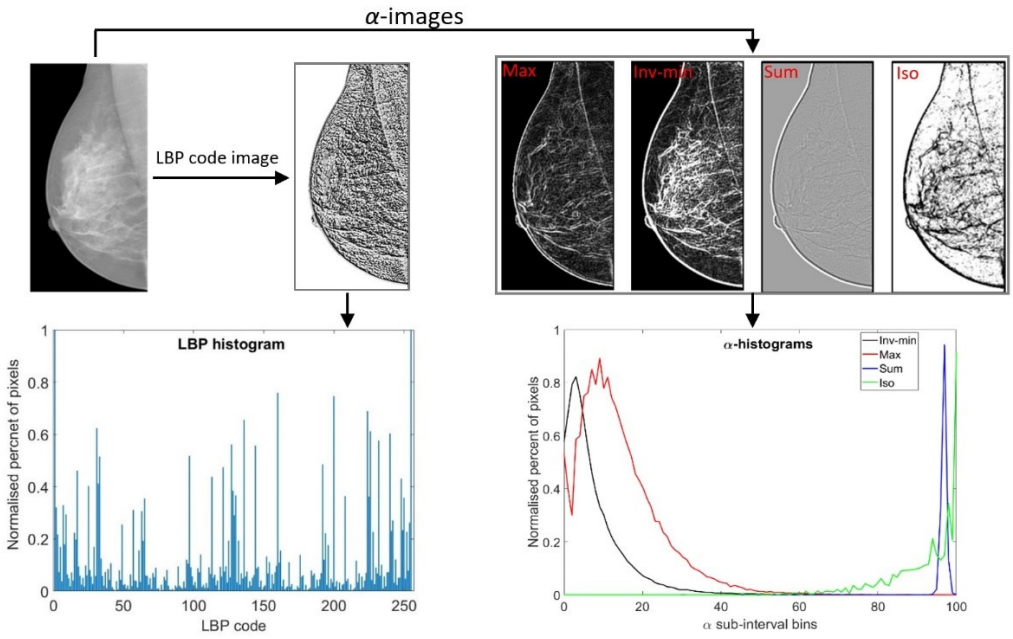


Figure 5.12: The normalised LBP histogram and α-histograms.

Table 5.4 Important parameters and their values used in the classification models.

Method	Related Parameters
LBP	1. ($R = 2$, $P = 8$)
MLBP	1. ($R_1 = 2$, $R_2 = 4$, $P = 8$)
ELBP	1. ($R_1 = 1$, $P_1 = 8$), ($R_2 = 4$, $P_2 = 8$)
Autoencoder network	1. The number of hidden layers tested: {5, 7, 9, 11, 13, and 15} 2. n: feature number
SVM classifier	1. Kernel: RBF 2. γ =10 3. C =1

5.3.5.1 Classification Results by Using Different LBP operators

As discussed in section 5.3.1, three different LBP operators are used and compared in the classification task. The parameter P is consistently set to 8, and 256 bins are generated in the histogram. For MLBP, two LBP histograms with R equals 1 and 2 are cascaded for including more local texture information; therefore, 512 bins are contained in its feature vector. As not all texture features in the generated histogram play the same role in the classification performance, PCA is used to produce a new feature vector in decreasing order and the top n features are tested. The three LBP based operators are

compared by analysing the accuracy of BI-RADS density classification. Fig. 5.13 shows the comparative results using different numbers (n) of selected features, which indicates that LBP produces a slightly better classification result (76.92%) when the feature vector length (n) is reduced to 35.

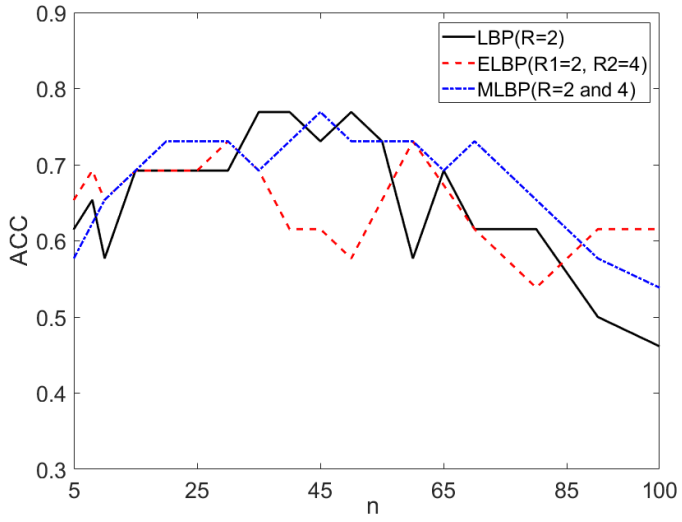


Figure 5.13: Classification accuracy comparison among different LBP based operators.

5.3.5.2 Classification Results by Using Cascaded Features

As shown earlier, α images in multifractal analysis provide a significant enhancement of texture patterns related to fibro-glandular tissue areas. The LBP-histogram is therefore concatenated with different α-histograms for capturing more effective image features and improving the classification performance. Each feature vector contains 356 features obtained by concatenating a 256-bin LBP histogram and a 100-bin α-histogram. An Autoencoder network and PCA techniques are used to optimise the feature vector with the aim of producing an n-dimensional feature vector used for classification. Different number of hidden layers in the autoencoder structure are tested in a set of {5, 7, 9, 11, 13, and 15} and an 11-layer architecture (as shown in Fig. 5.11) with the code layer size s = 64 is used in experiments. When using the autoencoder network, the number of neurons in each hidden layer is set to constantly decrease from its input layer to the core code layer for the encoder part, and the decoder part uses the inverse operations. PCA is applied on the reduced 64-length feature vector, outputting a new feature set based on which the top n features are tested.

Fig. 5.14 shows that the feature set containing LBP and α-Iso histograms (LBP-α(Iso)) reaches a higher classification accuracy (80.77%) than the other combined

feature sets, and only 14 extracted features are used in this classification model. The experimental results also demonstrate that the use of the autoencoder network reduces the feature dimensionality effectively by compressing the feature vector from 356 features to 14, while maintaining a desirable classification accuracy.

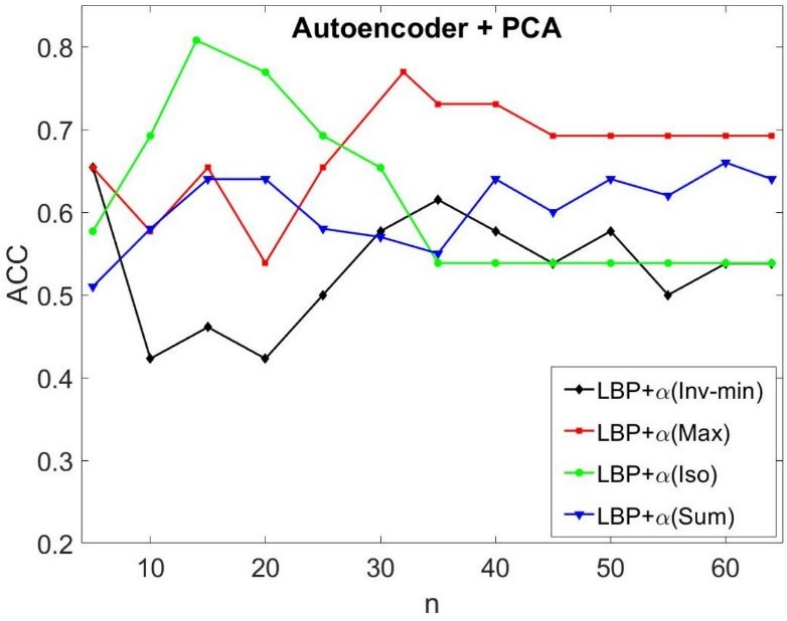


Figure 5.14: Autoencoder network and PCA are used to reduce the feature dimensionality and the cascaded feature set of LBP-α (Iso) reaches the best classification accuracy.

5.3.5.3 Results Comparison

This section investigates texture features extracted from LBP and α-histograms for classifying breast density, and different combinations of texture feature sets are tested and compared. In a recent study (Rampun et al., 2018), a local quinary pattern (LQP) method is applied to extract texture features using different neighbourhood topologies with the same classification task. Their results indicate that the ellipse topology based LQP gives the best accuracy of 82.02% when using over 200 image features to test 206 images in the INbreast dataset (only MLO view images used). By contrast, our proposed classification model is tested on the whole INbreast dataset (409 images) and reaches accuracy of 80.77% using only 14 features. Table 5.5 gives the comparison of classification performance by using LQP, LBP and the cascaded texture feature sets.

Table 5.5 The classification accuracy comparison between different texture features.

Image feature	The number of features used	The number of test images	Classification accuracy (%)
LQP(Ellipse)	Over 200	206	**82.02**
LQP(Circle)	Around 100	206	Under 80.00
LBP	35	409	76.92
ELBP	30	409	73.08
MLBP	45	409	76.92
LBP-α(Inv-min)	**5**	409	65.40
LBP-α(Max)	32	409	76.92
LBP-α(Iso)	**14**	409	**80.77**
LBP-α(Sum)	60	409	66.00

5.4 Chapter Summary

This chapter introduces two breast density classification models. The first model uses the multifractal spectrum as a texture feature to classify mammogram patches to dense or fatty tissue categories. Histograms of pixel intensity in the breast region are analysed to develop adaptive alpha value selection rules for producing more effective multifractal spectra. Using this method, dense tissue areas could be recognized in the mammogram image and percentage density calculated to match BI-RADS category. The second classification model concatenates LBP and alpha histogram information to represent more texture features related to fibro-glandular tissue. To optimise the extracted feature vector and reduce the feature dimensionality, this model adopts an autoencoder network and principal components analysis method to process the initial feature set. The INbreast dataset containing 409 mammogram images with four BI-RADS density categories is used to test the two classification models. Experimental results demonstrate that the use of multiple image features in the proposed models can improve the capability of image representation and produce desirable classification results for evaluating breast density.

6 Robust Texture Features for Breast Density Classification in Mammograms

This chapter presents a breast density classification model with an effective and robust texture feature descriptor. Local quinary pattern (LQP) method is considered and improved in this work for developing a richer set of texture features. A rotation invariant approach with different numbers of spatial bit-transitions is used to extend LQP to rotation invariant uniform LQP (RIU4-LQP). The proposed feature descriptor recognizes more texture patterns and reduces high feature dimensionality significantly. In addition, this chapter investigates the impact of using resized mammogram images on breast density classification results. Two mammogram datasets, INBreast and MIAS, are used in our experiments to test the proposed model. Compared to state-of-the-art methods, competitive classification results are obtained, with classification accuracies of 82.50% and 80.30% on INBreast and MIAS datasets, respectively. Comparative statistical analysis indicates that the proposed method outperforms other approaches.

6.1 Proposed Classification Model

Several recent research work have extended LBP to different variants to describe richer image representations. (Tan & Triggs, 2007) modified the approach and introduced Local Ternary Patterns (LTP), which threshold the neighbouring pixels using a three-value encoding system based on a constant threshold set by the user. (Nanni, Luminia, & Brahnam, 2010) introduced a five-value encoding system called Local Quinary Patterns (LQP). In other words, LBP, LTP and LQP threshold the neighbouring pixels into two (1 and 0), three (−1, 0 and 1) and five (2, 1, 0, −1 and −2) values, respectively. Related work shows that improved texture descriptors based on LBP and LQP can be used to classify breast density effectively, with promising classification results (Rampun, Scotney, Morrow, Wang, & Winder, 2018) (George & Zwiggelaar, 2019). However, in order to include more local region information, the parameters R and P in these methods are set to larger values which lead to an exponential growth in the number of image features. In such cases, the curse of high feature dimensionality is a major limitation (Yelampalli, Nayak, & Gaidhane, 2019) for classifying target mammograms into a limited number (3 or 4) density categories. Considering the above aspects, this chapter extends the LQP operator to a new texture feature descriptor RIU4-LQP using the rotation invariant approach. This extension work has two aims: i) to reduce the high dimensionality of features (number of bins of the histogram); ii) to increase the effectiveness of representing tissue structures in mammograms using texture features. The classification model based on RIU4-LQP is shown in Fig. 6.1, and more details are presented in the following sections.

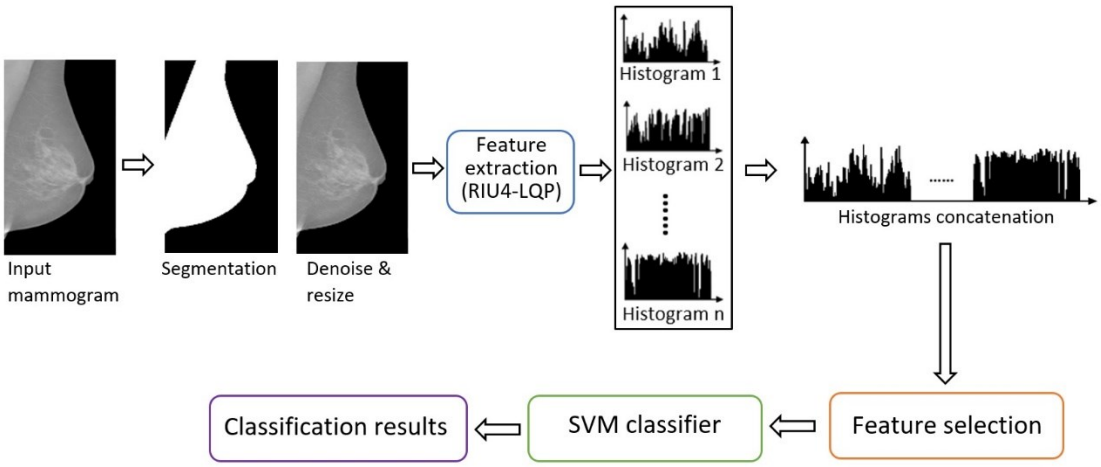


Figure 6.1: An overview of the classification model.

6.2 Texture Feature Extraction Method

This section first introduces the process of extracting texture features in mammograms using the basic LQP operator. Subsequently, a rotation invariant method is applied in LQP to reduce the feature numbers. In order to avoid compressing those important and discriminative texture features when using rotation invariance, this work also investigates the use of different transition number conditions to increase the number of invariants. A new texture feature descriptor RIU4-LQP is developed and used in the proposed classification model, and experimental results in section 6.4 demonstrate the effectiveness and robustness of this novel texture feature descriptor.

6.2.1 Feature Extraction Using LQP

(Nanni, Lumini, & Brahnam, 2010) proposed LQP by extending LBP from a binary value encoding scheme to a 5-value encoding algorithm. LQP uses 5 values (−2, −1, 0, 1, and 2) to describe relations between the intensity value of a central point and its neighbours. As shown in Fig. 6.2, each LQP code can be split into 4 LBP patterns by using Eqs. (6.2) − (6.5), capturing more detailed texture information. Therefore, for analysing medical images such as mammograms which represent subtle texture differences, texture features extracted by LQP contain more information useful for improving the final classification accuracy. To implement the LQP operator, two threshold values $\{\tau_1, \tau_2\}$ are required for generating a 5-value encoding pattern. The calculation of the LQP code can be described as follows:

$$LQP_j(c,R,P)=\sum_{i=0}^{P-1}s_j(g_i-g_c)2^i, \quad j=1,2,3,4 \tag{6.1}$$

$$s_1(x)=\begin{cases}1, & if\ d(x)=2\\ 0, & otherwise\end{cases} \tag{6.2}$$

$$s_2(x)=\begin{cases}1, & if\ d(x)=1\\ 0, & otherwise\end{cases} \tag{6.3}$$

$$s_3(x)=\begin{cases}1, & if\ d(x)=-1\\ 0, & otherwise\end{cases} \tag{6.4}$$

$$s_4(x)=\begin{cases}1, & if\ d(x)=-2\\ 0, & otherwise\end{cases} \tag{6.5}$$

$$d(x)=\begin{cases}2, & \tau_2 \le x \\ 1, & \tau_1 \le x < \tau_2 \\ 0, & -\tau_1 \le x < \tau_1 \\ -1, & -\tau_2 \le x < -\tau_1 \\ -2, & x < -\tau_2\end{cases} \tag{6.6}$$

where R denotes the radius used to specify a circular neighbourhood and P is the number of neighbourhood pixels used to calculate the LQP code. g_i and g_c are intensity values of the pth neighbour pixel and the centre pixel c, respectively. From a specific position (usually the top-left corner) the binary values given by $s_j(x)$ are gathered in a specific sequence (usually in a clockwise order) to obtain the LQP_i codes. Fig. 6.2 illustrates how the LQP values are computed in a local region with $R = 2$, $P = 16$, using a threshold value set $\{\tau_1 = 2, \tau_2 = 7\}$. After obtaining the 5-value pattern, it is split into 4 binary patterns by $s_j(x)$ in Eqs. (6.2) ─ (6.5).

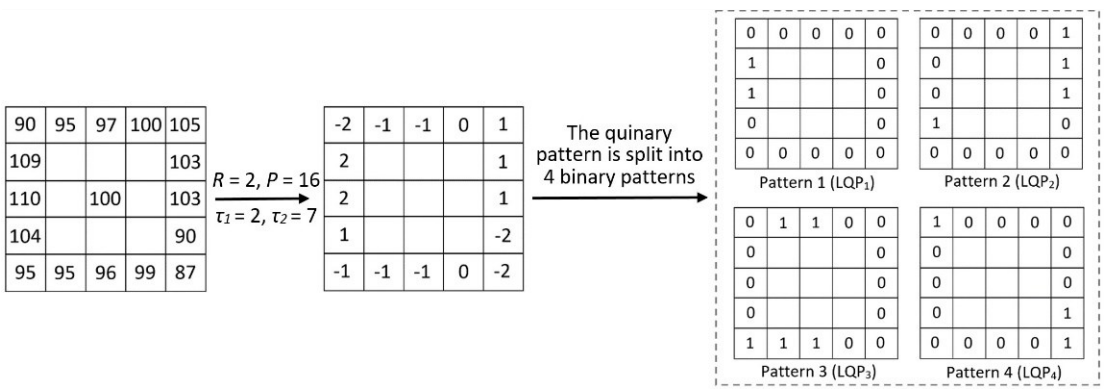


Figure 6.2: LQP encoding example, one 5-value pattern is split into 4 LBP patterns.

Following the calculation of the LQP codes, four texture pattern images (LQP_i images) can be generated by using the LQP_i code to replace the original intensity value of point c. LQP_i histograms are commonly used as image features to characterise texture patterns. As shown in Fig. 6.3, four histograms of LQP_i images are concatenated to form the final feature vector.

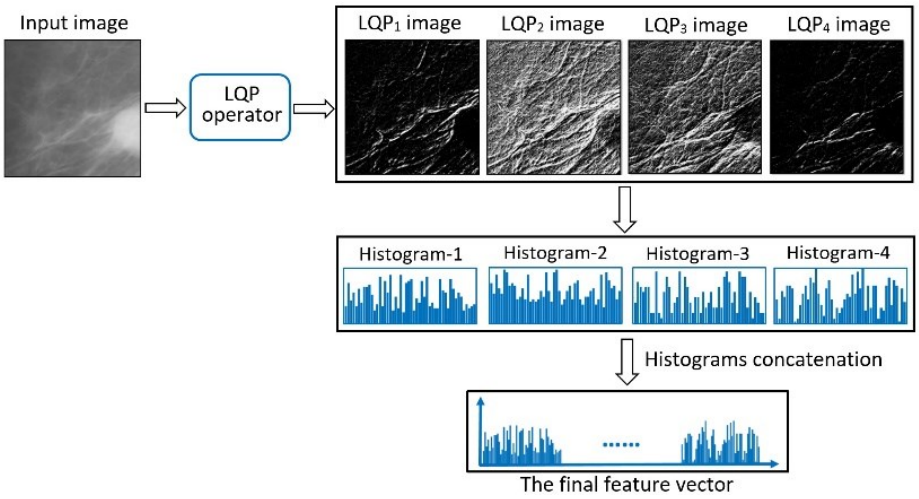


Figure 6.3: Example of outputting LQP based feature vector within a local region of mammogram.

As shown in Fig. 6.2 and Fig. 6.3, the LQP feature vector can have the problem of high feature dimensionality. We can see that the feature dimensionality increases exponentially by 2^P in LBP and 2^{P+2} in LQP. A large feature space cannot be utilised efficiently to train a classification model, and information redundancy in the feature set can have a negative impact on the final classification performance.

6.2.2 Rotation Invariant Method

LQP features are grey-scale invariant, but not rotation invariant. The same texture patterns with different rotation angles could correspond to different pattern codes. We first use a rotation invariant strategy (Guo, Zhang, Zhang, & Zhang, 2010) to extend LQP to rotation invariant uniform LQP (RIU2-LQP) as follows:

$$LQP_j^{riu2}(c,R,P)=\begin{cases}\sum_{i=0}^{P-1}s_j(g_i-g_c), & if\ T_j(LQP(c,R,P))\in\{0,2\}\\ P+1, & otherwise\end{cases} \tag{6.7}$$

$$T_j(LQP(c,R,P))=\left|s_j(g_{P-1}-g_c)-s_j(g_0-g_c)\right|+\sum_{i=1}^{P-1}\left|s_j(g_i-g_c)-s_j(g_{i-1}-g_c)\right| \tag{6.8}$$

where, $j \in \{1, 2, 3, 4\}$; $T_j(\cdot)$ is defined as the number of spatial transitions (bitwise 0/1 changes) in patterns.

By using the extended RIU2-LQP, we can see that the extracted feature dimensionality is reduced significantly. For example, when $P = 12$, the common LQP method produces a histogram of length 16384 bins ($2^{12} \times 4$) after concatenating the histograms of the four patterns, while RIU2-LQP produces a 56-bin histogram (14×4).

However, such dimensionality reduction may suppress important texture information which may in turn affect the final classification performance. Fig. 6.4 shows 4 different texture patterns (a) ─ (d) when R = 4, P = 16, and RIU2-LQP only recognizes the first texture pattern (LQP^{riu2} = 8), as the other 3 patterns (b) ─ (d) are labelled with the same pattern code (LQP^{riu2} = 17).

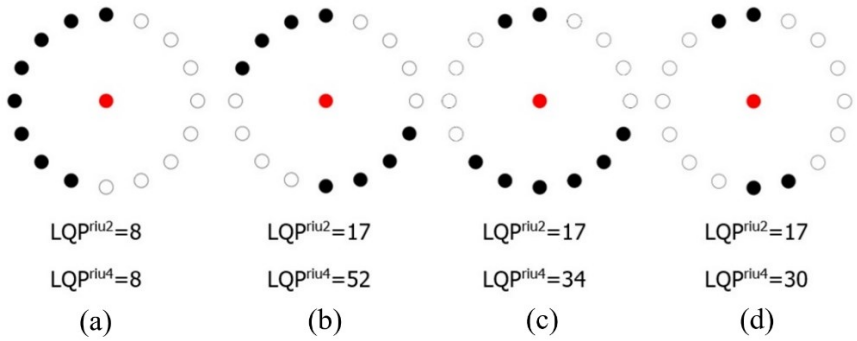


Figure 6.4: Examples of different texture patterns and their LQP codes when P = 16 (black circle: 1, empty circle: 0).

6.2.3 Uniform Coding Method Using a Higher Number of Bit Transitions

Based on the drawback of RIU2-LQP, this study further extends it to RIU4-LQP by analysing the transition number in a wider range, i.e. $T \in \{0, 2, \text{and } 4\}$. Compared with RIU2-LQP, which considers only one '1'-contiguous segment in each extracted binary pattern (e.g. Fig. 6.4 (a)) and ascribes all other patterns to one group, RIU4-LQP is extended to analyse patterns with two '1'-contiguous segments. In this way, RIU4-LQP can capture richer image representations of texture patterns. The proposed RIU4-LQP encoding scheme can be formulated by the following equations.

$$LQP_j^{riu4}(c,R,P) = \begin{cases} \sum_{i=0}^{P-1} s_j(g_i - g_c), & if\ T_j(LQP(c,R,P)) \in \{0,2\} \\ index + P, & if\ T_j(LQP(c,R,P)) = 4 \\ \left\lceil \frac{P^2+11}{4} \right\rceil - 1, & otherwise \end{cases} \tag{6.9}$$

$$index = \begin{cases} \sum_{n=1}^{X-1} (P-3-2(n-1)) + (Y-X+1), & if\ X \geq 2 \\ Y-X+1, & if\ X = 1 \end{cases} \tag{6.10}$$

where, X and Y denote the number of occurrences of '1' in two distinct contiguous segments when $T = 4$ (e.g. Fig. 6.4 (b) ─ (d)). We impose a restriction with respect to the relation between X and Y: X always denotes the shorter '1'-contiguous segment, i.e. $X \leq Y$. For example, in Fig. 6.4 (c), $X = 2$, $Y = 6$. Eq. (6.9) contains three different parts corresponding to numbers of bit transitions (i.e. the value of T) for recognizing and encoding different texture patterns. The first part (i.e. the first row of Eq. (6.9)) is the same as Eq. (6.7) which generates RIU2-LQP code for distinguishing texture patterns that have none or only one '1'-contiguous segment (e.g. Fig. 6.4 (a)). The second part of Eq. (6.9) aims to encode texture patterns with two '1'-contiguous segments (e.g. Fig. 6.4 (b) ─ (d)). As the encoding values from 0 to P has been allocated in the first part (i.e. $T \in \{0, 2\}$), the output code starts from P and adds the value of *index* which counts from 1 using Eq. (6.10). The value of *index* is used to recognize and label those texture patterns when $T = 4$, for example, *index* = 1 with the pattern of $X = 1$, $Y = 1$; *index* = 2 with the pattern of $X = 1$, $Y = 2$; and so on. A detailed description of Eqs. (6.9) and (6.10) is given in appendix 1. The third part of Eq. (6.9) uses a unified code to denote all the other texture patterns presenting three or more '1'-contiguous segments. By using the proposed RIU4-LQP encoding method, LQP^{riu4} codes are allocated to different texture patterns in Fig. 6.4.

Comparing with the conventional LQP method, the extended RIU4-LQP reduces the feature dimensionality from $2^P + 2$ to $P^2 + 11$. Meanwhile, compared to RIU2-LQP, more texture patterns are included in the extracted features by considering a higher number of bit transitions.

6.3 Feature Selection Method and Classifier

After computing image texture features, this model conducts a feature selection process using a dominant pattern selection method to remove the redundant texture features. In the initial feature set which consists of multiple histograms of RIU4-LQP in different pattern channels, not all texture features play an equally important role in the classification work. Therefore, an appropriate feature selection step, filtering the most important image features in the final feature set, can improve the classification performance. In addition, related work in the literature (Tzikopoulos, Mavroforakis, Georgiou, Dimitropoulos, & Theodoridis, 2011)(Oliver et al., 2015) (Rampun, Scotney, Morrow, Wang, & Winder, 2018) (Rampun, Morrow, Scotney, & Wang, 2020) reported

their best classification results by using SVM to predict density labels for mammograms. Therefore, SVM is used in this proposed model and is trained using the extracted texture feature set.

6.3.1 Dominant Pattern Selection

A dominant pattern selection (DPS) method proposed by Guo et al. (Guo, Zhao, & Pietikäinen, 2012) is used in this classification model to select image features. A set of dominant patterns for an image is the minimum set of pattern types that can cover n% ($0 < n < 100$) of all patterns of an image. Therefore, a bin-wise summation for all the histograms in a training set is performed to output a resulting histogram (H) with M bins. The distribution of dominant patterns (bins of the histogram) can be observed from it. Then the histogram (H) is sorted in descending order, and the top N bins are selected as follows:

$$\arg\min_{N} \frac{\sum_{i=1}^{N} H(i)}{\sum_{j=1}^{M} H(j)} \geq n\% \tag{6.11}$$

where M is the total number of bins of H; n is a threshold value for the percentage that the selected dominant patterns account for in H.

6.3.2 Classification

To maintain consistency and uniformity in the evaluation of the classification algorithm, we use the same SVM classifier for each considered descriptor with 10-fold cross validation. For SVM classifier, three different kernels, RBF, Poly and Sigmoid, are used and compared for pursuing the optimal classification accuracy (CA). In this model, poly kernel is finally used, as better classification results are observed (shown in Fig. 6.5). Other parameters (gamma, C and degree) are found using a grid-searching technique.

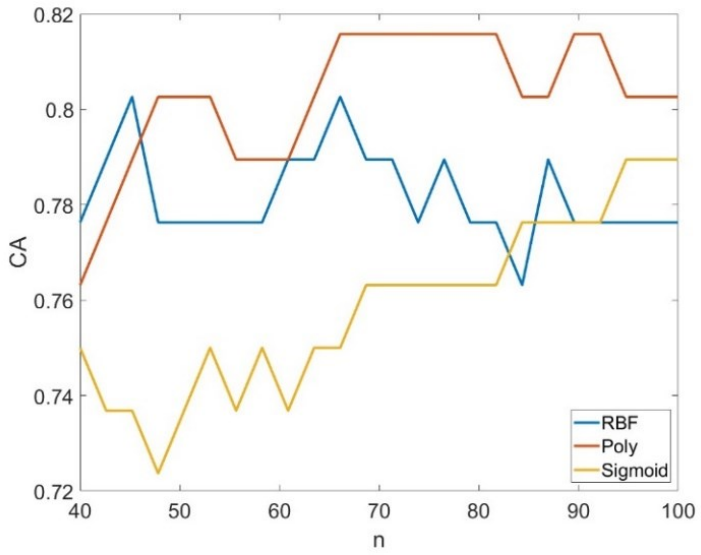


Figure 6.5: Testing different kernels of SVM classifier using RIU4-LQP.

6.4 Experimental Results and Comparative Analysis

To demonstrate the effectiveness of the proposed texture feature descriptor used in the classification model, different texture feature extraction methods based on LBP/LQP are tested and compared in experiments. In addition, mammogram images are resized by a scaling factor (s) to see the difference on classification results. Two mammogram datasets, INbreast and MIAS, which contain 409 and 322 images, respectively, are used to test the proposed classification model. Experimental results are evaluated by classification accuracy (CA) and area under the ROC curve (AUC). A statistical test (t-test) is conducted to see the difference between compared methods.

6.4.1 Compared Methods

As the proposed RIU4-LQP method is an extended feature descriptor based on the LBP/LQP method, we design the following 5 progressive transformations of LBP/LQP and implement the corresponding algorithms to classify mammograms into different density categories. Comparative analysis is done based on their classification results.

1) *LBP:* The basic LBP method with multi-orientation and multi-resolution approaches used for extracting LBP codes. Specifically, a combination of three different R values (i.e. R = 2, 4, 8) is used at the feature extraction step in experiments.

2) *RILBP:* The common rotation invariant method is applied when computing LBP codes.

3) *LQP:* The multi-orientation and multi-resolution methods are used on the basic LQP method. The threshold set $\{\tau_1, \tau_2\}$ in LQP are selected and tested from the set of {1, 2, 3, 4, 5, 6, 7, and 8}.

4) *RIU2-LQP:* The extended LQP method with rotation invariant features. Three pairs of (R, P) (R = 2, 4, 8; P = 10, 14, 18) are used to calculate texture pattern codes.

5) *RIU4-LQP:* The proposed rotation invariant uniform LQP with $T_j \in$ {0, 2, 4, and others}. The same multi-resolution configuration in 4) is applied for its computation.

Detailed parameters and their values used in the experiments are given in Table 6.1.

Table 6.1 Relative parameters and theirs values used in the experiments.

Method	Related Parameters
Image pre-processing	1. Resizing scale s = {1/16, 1/8, 1/4, 1} 2. Median filter size: 3 × 3
LQP	1. R = 4, P = 8 2. Multi-orientation method considers three orientation angles: 0°, 15°, 30°
RIU2-LQP RIU4-LQP	In multi-scale method, three pairs of (R, P) are used: 1. (R_1 = 2, P_1 = 10), (R_2 = 4, P_2 = 14), (R_3 = 8, P_3 = 18) 2. Threshold values $\{\tau_1, \tau_2\}$ are decided by an adaptive method
Features selection	Dominant pattern selection: 1. Dominant patterns covering percentage (n%): 40% ~ 100%
SVM classifier	1. Kernel: RBF 2. γ =10 3. C =1

6.4.2 Comparison Using Different Scaling Factors

Different values of the scaling factor (s) are evaluated to reveal the difference on classifying results by using resized mammograms. The value of s is set to decrease from 1 to 1/16. Fig. 6.6 shows the classification accuracies obtained using resized mammograms in the classification model. A higher classification accuracy is seen when s = 1/8 on INbreast.

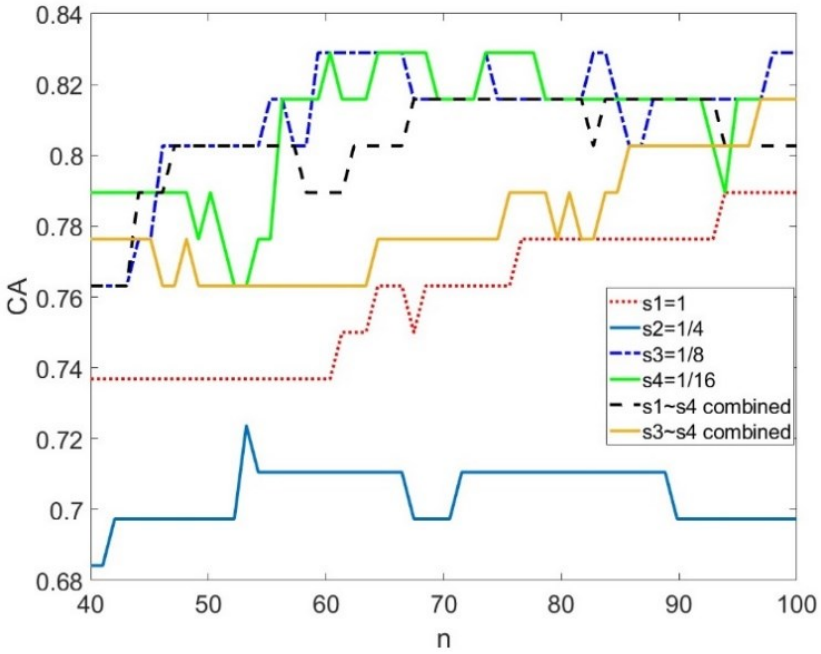


Figure 6.6: Classification accuracy comparison using RIU4-LQP method and differently resized mammograms.

6.4.3 Classification Performance Evaluation

As this is a multi-category classification task, AUC is calculated by the mean micro-average ROC method on 10-fold cross validation. A statistical test is done in order to ensure that experimental results are not discovered by coincidence. The "10-fold cv t-test" method in (Bouckaert, 2003) with a significance level of 0.05 (i.e. *alpha* = 0.05) is used for RIU4-LQP and every other method to calculate a *p*-value, which shows the statistical difference between methods. The parameter *n* in Eq. (6.11) is compared as well.

Table 6.2 and Table 6.3 show the classification results on two different datasets. We can see that the proposed method (RIU4-LQP) outperforms other methods by obtaining higher CA and AUC values on both mammogram datasets. The highest classification accuracy is 82.50 ± 8.75 on INbreast and 80.30 ± 4.55 on MIAS. Except for LQP (p = 0.205), other methods present low p values (< 0.05), which means that the difference on the classification performance between RIU4-LQP and other methods is statistically significant.

Table 6.2 Classification results on INBreast are evaluated by classification accuracy (CA), area under the ROC curve (AUC) and 10-fold cv t-test.

	CA	AUC	Best n	*p*-value
LBP	66.25 ± 10.16	79.72 ± 8.54	65.7	0.002
RILBP	70.00 ± 8.29	84.95 ± 6.47	99.4	0.003
LQP	80.63 ± 10.62	88.32 ± 8.37	79.1	0.205
RIU2-LQP	77.50 ± 9.35	88.16 ± 8.73	96.8	0.028
RIU4-LQP	**82.50 ± 8.75**	**89.03 ± 7.61**	96.7	-

Table 6.3 Classification results on MIAS are evaluated by classification accuracy (CA), area under the ROC curve (AUC) and 10-fold cv t-test.

	CA	AUC	Best n	*p*-value
LBP	74.55 ± 10.77	85.61 ± 7.72	71.8	0.045
RILBP	67.58 ± 10.68	88.78 ± 5.96	83.2	0.001
LQP	76.06 ± 8.72	87.51 ± 6.45	66.7	0.062
RIU2-LQP	74.24 ± 8.60	86.46 ± 6.86	81.7	0.014
RIU4-LQP	**80.30 ± 4.55**	**88.87 ± 3.56**	81.5	-

6.4.4 Comparison Between Two-view Groups

In the INBreast dataset, each subject is associated with two different views (MLO and CC) of mammograms. INbreast mammograms are divided into MLO or CC view groups along with the subject ID information (as shown in Fig. 6.7), based on which classification results are analysed by paired t-test to show the statistical difference on the two different groups. With such pairwise division, we assume that the average classification performance on the two groups should be similar when using a texture descriptor and classification model. Using the proposed model, each test mammogram's category is predicted to be one of {0, 1, 2, and 3} which match to BI-RADS I-IV, respectively. The classification performance is evaluated by the predicted category value (mean ± stand deviation), classification accuracy and the *p*-value in the t-test (Table 6.4). The calculation of the *p*-value is based on the classification accuracy. A high *p*-value (e.g. $p = 1$ for RIU4-LQP) indicates that the predicted density results on the two view groups are statistically the same, which demonstrates that the proposed RIU4-LQP not only produces the best classification result, but also works equally well on MLO/CC view mammograms.

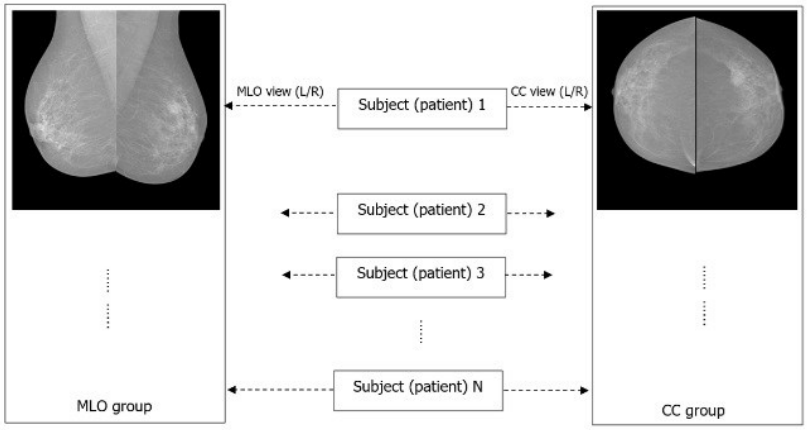


Figure 6.7: Mammograms in INBreast dataset are divided into two groups (MLO and CC views).

Table 6.4 Classification performance and t-test on INbreast MLO/CC view groups using different methods.

	Predicted category		Classification accuracy(CA%)		*p*-value
	MLO group	CC group	MLO group	CC group	
LBP	1.47 ± 1.01	1.53 ± 1.05	65.26 ±10.04	64.21 ± 8.74	0.619
RILBP	1.81 ± 1.12	1.73 ± 1.03	61.58 ± 6.25	74.21 ± 8.31	0.014
LQP	1.57 ± 1.04	1.83 ± 0.98	74.21 ± 6.42	65.26 ± 7.14	0.009
RIU2-LQP	1.88 ± 0.91	1.88 ± 0.85	75.65 ± 5.22	76.09 ± 7.08	0.758
RIL4-LQP	1.94 ± 0.89	1.88 ± 0.85	**80.53 ± 7.08**	**80.53 ± 5.79**	**1.000**

6.4.5 Running Time Comparison

This section compares running time cost using the proposed classification model. Note that the time cost is recorded for the 10 run 10-fold cross validation in which one fold of images is used as the test set and the remaining images are used to train the classifier for each iteration. The total time is divided by the number of images in a dataset, and we obtain the average time for processing one image, which covers training and classification procedures. We compare the time cost by using three different feature sets that are extracted with different R and P values. Table 6.5 gives the time cost comparison among LQP, RIU2-LQP and RIU4-LQP. We can notice that the running time of using LQP features increases significantly as the parameter P grows, and the feature dimensionality rises exponentially (introduced in section 6.2). By contrast, the time cost does not change a lot when using the RIU2-LQP and RIU4-LQP methods, remaining around 17~19 milliseconds per image. Particularly, when P is set to 18 in the

last column of Table 6.5 (i.e. considering more neighbouring points), the use of LQP costs more than 60 times as much as the other two methods on the INbreast, and around 50 times higher on MIAS. Programs implementing different methods are run on a desktop with Intel core i7 3.6GHz CPU, 16 GB memory, and Matlab R2017b. By considering the experimental results in Table 6.2−6.5, we can see that the proposed classification model with this novel texture descriptor produces the highest classification accuracy using less running time, which demonstrates its effectiveness and robustness.

Table 6.5 Time cost for training and classifying procedures by using different methods.

		Average time of classifying image (milliseconds per image)		
	Methods	**R = 2, P = 10**	**R = 4, P = 14**	**R = 8, P = 18**
INbreast	**LQP**	28.66	241.24	1220.82
	RIU2-LQP	17.45	17.76	17.91
	RIU4-LQP	18.19	19.09	19.89
MIAS	**LQP**	25.95	197.22	913.40
	RIU2-LQP	17.43	17.65	17.75
	RIU4-LQP	17.67	18.87	19.31

6.5 Chapter Summary

In this chapter, we develop an effective and robust texture feature descriptor for classifying mammographic breast density. Based on the commonly used local binary patterns (LBP), this study investigates its variant method local quinary patterns (LQP) and extends it to rotation invariant uniform LQP (RIU4-LQP), which reduces the dimensionality of its feature vector significantly. For recognizing diverse texture patterns, the rotation invariant approach is extended by considering different numbers of spatial transitions and is applied in the LQP method. The new RIU4-LQP method is used to extract robust texture features in mammograms with the aim of classifying breast density into different categories. Two mammogram datasets (INBreast and MIAS) with different classification criteria are used to test the proposed classification model. Experimental results show that the use of the RIU4-LQP feature descriptor produces promising classification accuracies of 82.50% and 80.30% on INbreast and

MIAS, respectively. Comparative analysis among different methods indicates that the classification performance using RIU4-LQP statistically outperforms that of other methods.

7 SPATIAL DISTRIBUTION ANALYSIS OF TEXTURE DESCRIPTORS

This chapter addresses a new breast density classification model combining image texture features and their corresponding spatial characteristics. The rotation invariant uniform local quinary pattern (RIU4-LQP) method proposed in the previous chapter is used to describe texture patterns in mammograms. In conventional processing methods, image features are obtained by computing histograms from the texture patterns. However, such processes ignore very important spatial information related to the texture features. This model explores richer statistical information extracted from the RIU4-LQP feature set by using Baddeley's K-inhom method to characterise spatial distribution information of the feature point sets. A new feature vector called 'K-spectrum', which offers supplementary and important image features, is developed based on the spatial characteristics. Histograms and K-spectra are first extracted and tested separately and then are combined to generate a new texture feature vector. A novel feature selection step is considered carefully in this study for removing redundant image features and improving the classification performance. The main contributions in this chapter are summarised as follows.

1) Based on the proposed texture descriptor RIU4-LQP, this study analyses its spatial distribution characteristics using Baddeley's K-inhom function, constructing a new feature vector called 'K-spectrum'.
2) A hybrid texture feature vector concatenating histograms and K-spectra is proposed and used to classify breast density in mammograms.
3) Machine learning based feature selection methods are employed to optimise the extracted feature set.
4) Comparative analysis and statistical tests are conducted on two publicly available mammogram datasets, which demonstrate the superiority of the proposed classification model.

In our experiment, two mammogram datasets, INBreast and MIAS, are used to test the proposed methods for classifying mammograms into different density categories. Classification performance is analysed based on the ground truth information annotated by breast imaging radiologists. Experimental results show that our proposed method outperforms other approaches in the literature and classification models in previous chapters, with the best classification accuracy of 92.76% (INbreast) and 86.96% (MIAS).

7.1 Proposed Classification Model

In the proposed model, the pre-processing steps introduced in Chapter 3 are first used to segment breast regions, remove noise and resize the original images. A novel feature descriptor RIU4-LQP introduced in Chapter 6 is used to produce RIU4-LQP images in different pattern channels and collect corresponding texture features. Based on the RIU4-LQP images in which texture patterns are enhanced, Baddeley's K-inhom method is employed to characterise spatial distribution information and construct a new feature vector 'K-spectrum'. Subsequently, RIU4-LQP based histogram and K-spectrum information are concatenated to form a new feature space which is used to classify mammographic density in this work. An overview of the workflow of the classification model is shown in Fig. 7.1, and more details of the methodology are given in the following sections.

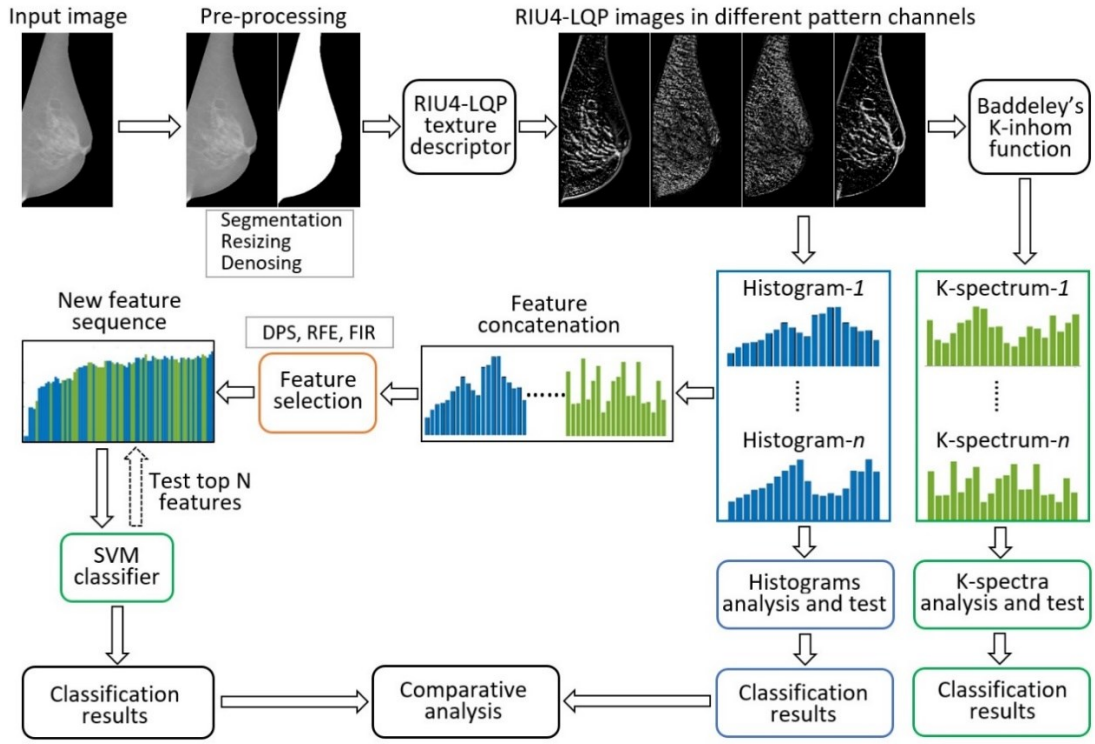


Figure 7.1: An overview of classifying breast density using the proposed methods.

7.1.1 Feature Extraction Using RIU4-LQP

As discussed in Chapter 6, the local quinary patterns (LQP) operator is extended by using a rotation invariant method and a wider range of transition number. With such an extension, the feature dimensionality is reduced significantly and the new encoding operator has a strong capability of recognizing typical texture patterns. Fig. 7.2 shows the comparison between RIU2-LQP and RIU4-LQP images in different pattern channels, from which we can see that RIU4-LQP images represent richer texture pattern information related to fibro-glandular tissue as they have a wider encoding range than RIU2-LQP ($P^2 + 11$ vs $4 \times (P + 2)$). In addition, for capturing more local texture features, a multiscale scheme including three pairs of (R, P) is applied in RIU4-LQP (shown in Fig. 7.3) in the proposed classification model. In conventional applications, image texture features are finally represented by histogram information which counts the quantity of feature points in histogram bins. Fig. 7.4 illustrates how one RIU4-LQP image is decomposed into pattern images in different encoding channels and a histogram is constructed based on it. However, the histogram reflects only the quantity information of related feature points, and not their spatial distribution information, which may be useful in distinguishing between dense or fatty tissue areas in breasts as shown in Fig. 7.4. We have not found any related work in the literature that used the

spatial distribution of texture feature points. In the following sections, we discuss a novel method to describe the spatial information and construct a new feature vector for the proposed classification model.

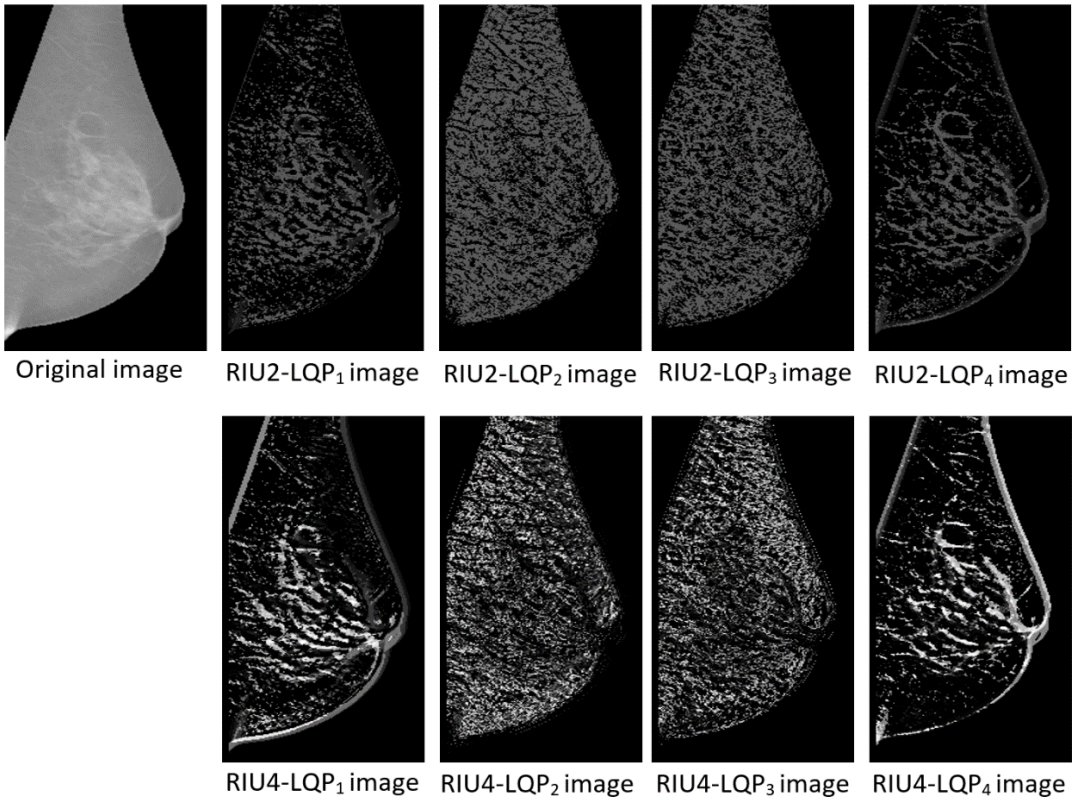


Figure 7.2: LQP images in different texture pattern channels when using RIU2-LQP and RIU4-LQP operators.

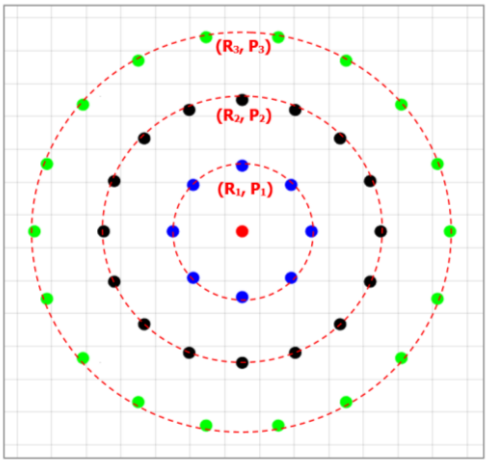


Figure 7.3: Multi-scale method with three pairs of (*R*, *P*) is used when conducting RIU4-LQP encoding system.

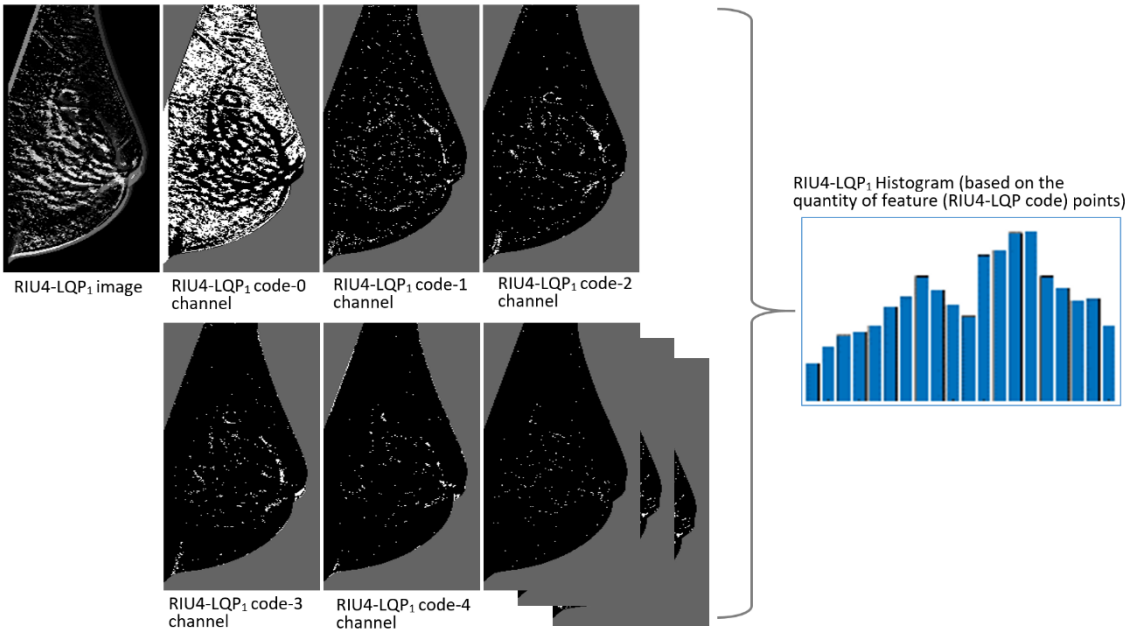


Figure 7.4: Decomposition of RIU4-LQP1 image in Fig. 7.2 based on each encoding channel. The positions of pixels in the considered RIU4-LQP1 code channel are marked by white colour in the black breast region background.

7.1.2 Baddeley's K-inhom Function

Baddeley's K-inhom method is employed in this model to extract spatial information based on feature points. The K-inhom method is a variant of Ripley's K function (Ripley, 1976), and both are statistical analysis schemes used for studying qualitative or quantitative characteristics of spatialized data. Generally, the K-inhom function for one-dimensional data can be described as follows:

$$K_{inhom}(r) = \frac{1}{D}\sum_{i}\sum_{j\neq i}\frac{1\{\|x_i - x_j\| \le r\}}{\hat{\lambda}(x_i)\hat{\lambda}(x_j)}c(x_i, x_j; r) \tag{7.1}$$

$$D = \frac{1}{|W|}\sum_{i}\frac{1}{\hat{\lambda}(x_i)} \tag{7.2}$$

$$c(x_i, x_j; r) = \frac{\|x_i > r\|}{\sum_{j}\frac{\|x_j > r\|}{\hat{\lambda}(x_j)}} \tag{7.3}$$

where, $1\{\|x_i - x_j\| \le r\}$ denotes an indicator which is worth 1 if the distance between point x_i and point x_j is less than or equal to r or 0 otherwise; $c(x_i; x_j; r)$ corresponds to the correction of edge effects and W to the region of interest. A function $c(x_i, x_j; r)$ is

implemented as in (Baddeley & Turner, 2000) for edge corrections. $\lambda(x_i)$ denotes an intensity function around point x_i, which is defined by the number of neighbouring points (x_j) expected in a small area with x_i in the centre, but $\lambda(x_i)$ is not known in practice. Instead, '$\hat{\lambda}(x_i)$' is used in Eq. 7.1-7.3 as the estimate of $\lambda(x_i)$, which is implemented by a non-parametric method (Baddeley, Møller, & Waagepetersen, 2000). K-inhom function, including the edge correction function c and the parameter of λ, has been implemented in the R package, 'spatstat', as a standard function (Baddeley, Adrian, Rubak, & Turner, 2015). This work uses 'spatstat' to calculate the corresponding K-inhom values for the processed mammogram images. Values of $K_{inhom}(r)$ are calculated with different distance measurements (r), which results in a K_{inhom} curve by connecting all the observed values in a $K_{inhom}(r)$-r plot (Fig. 7.5). In the literature (Jean-Michel, Eric, & Florence, 2018), an expected reference value $K_{pois}(r) = \pi r^2$ (the red dotted line in Fig. 7.5), obtained based on an inhomogeneous Poisson process, is used to compare with the observed $K_{inhom}(r)$ value. If the K_{inhom} curve is located below the reference curve (i.e. $K_{inhom}(r) < K_{pois}(r)$), it indicates that corresponding points scatter regularly in the region of interest. By contrast, if the K-inhom curve is located above the reference curve (i.e. $K_{inhom}(r) > K_{pois}(r)$), the distribution of points tends to be more aggregated. Therefore, the K_{inhom} curve can be used to describe the spatial distribution characteristics of a point set. As the work in this thesis focuses on mammogram image analysis, we use the segmented breast region as the region of interest W (shown in Fig. 7.6). The pixels decomposed by RIU4-LQP code values (Fig. 7.4) constitute feature point sets in different code channels, which produce their K_{inhom} curves separately and show corresponding spatial characteristics. Fig. 7.6 gives an example of using one feature point set to produce its K_{inhom} curve, and Fig. 7.7 shows the examples of applying the K-inhom functions in local regions.

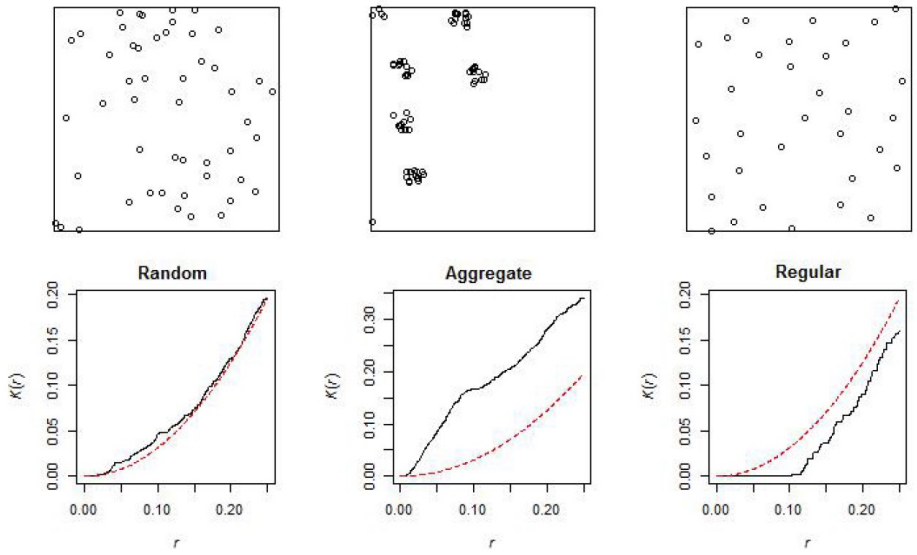


Figure 7.5: K_{inhom} curves of different point sets with distinct spatial distribution patterns (Baddeley & Turner, 2000).

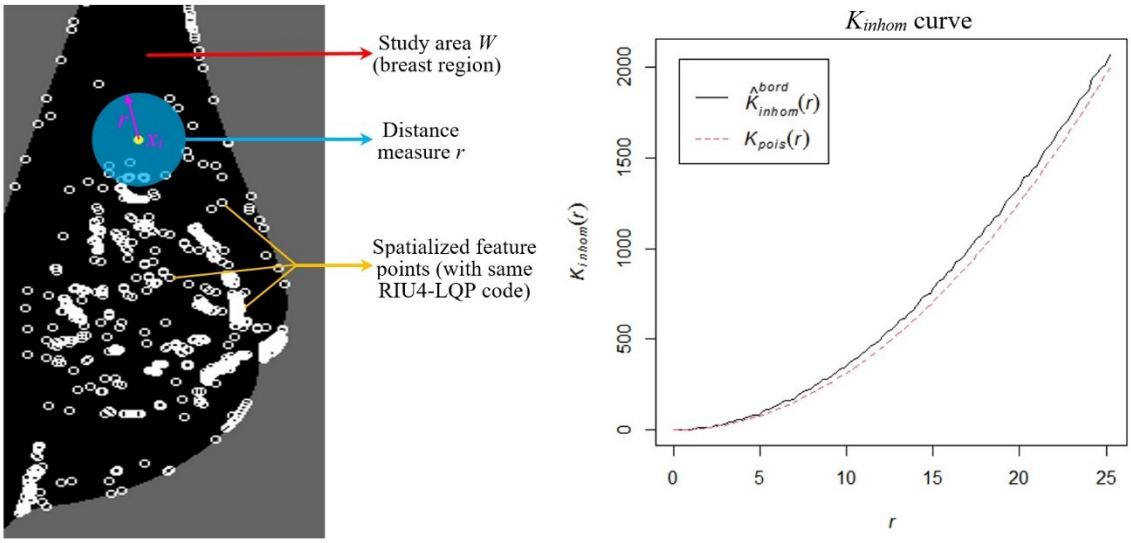


Figure 7.6: One feature pattern image (left) in a RIU4-LQP code channel with all features points are emphasized by white circles; the corresponding K_{inhom} curve (right) is generated by Baddeley's K-inhom function.

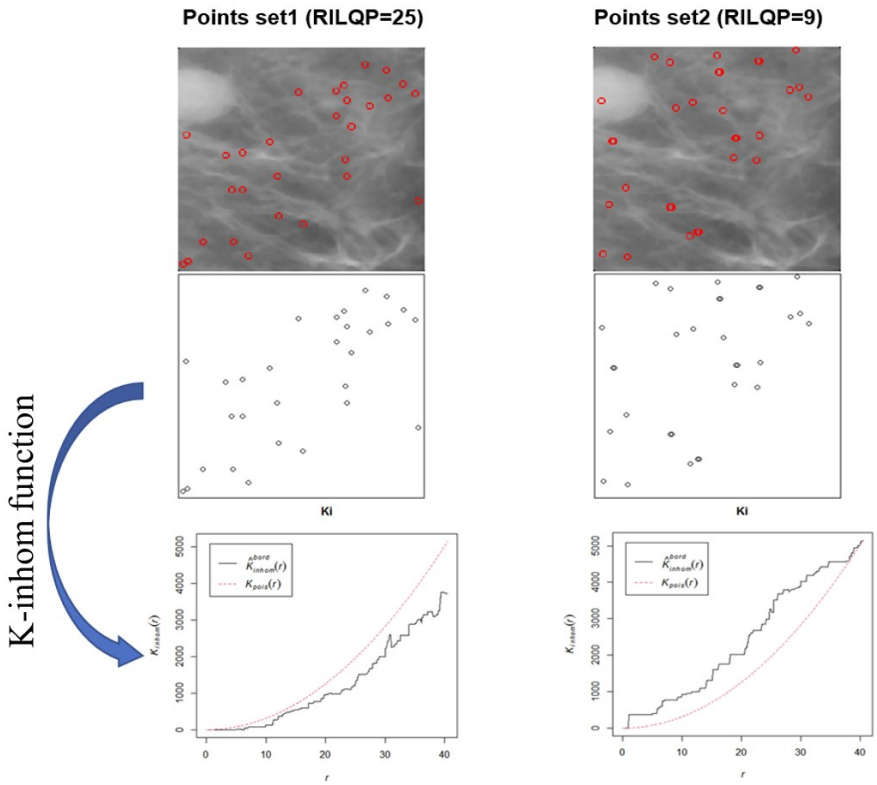


Figure 7.7: An example showing the application of K-inhom function in local regions of mammogram images.

7.1.3 K-spectrum of RIU4-LQP

This study employs Baddeley's K-inhom method (Baddeley, Adrian, Rubak, & Turner, 2015) to characterise spatial distributions of point sets with RIU4-LQP features, and a new texture feature vector called K_{inhom} curve based spectrum (K-spectrum) is generated with the aim of capturing extra texture information in mammograms. When using Eqs. (7.1) and (7.2) to produce K_{inhom} curves, x_i and x_j are two mammographic points belonging to one point set in which all the points have the same RIU4-LQP code value, and the region of interest W corresponds to the segmented breast region. The RIU4-LQP operator is used to produce an LQP_i^{riu4} code set $\{code\text{-}1, code\text{-}2 \ldots code\text{-}k\}$. Subsequently, pixels in the breast region are divided into k different point sets $\{X_1, X_2 \ldots X_k\}$ by corresponding LQP_i^{riu4} code values. Baddeley's K-inhom function is adopted to output a K_{inhom} curve for each point set X_i, which reflects how these points are scattered in the breast region with respect to a specific distance measure (r). As introduced in the last section, the reference curve $K_{pois}(r)$ is used for comparing with the observed K_{inhom} curve. Therefore, this study uses a deviation (d) between the observed $K_{inhom}(r)$ value and the reference value $K_{pois}(r)$ under the radius r and its mean ($\bar{d}$) on a valid r range to

evaluate the spatial distribution information of point sets. The deviation d and the mean $\bar{d}$ are computed as follows.

$$d(r) = K_{inhom}(r) - K_{pois}(r) \tag{7.4}$$

$$K_{pois}(r) = \pi r^2 \tag{7.5}$$

$$\overline{d_i} = \frac{\sum_{r=1}^{s} d_i(r)}{s},\ i = 1, 2, \ldots k. \tag{7.6}$$

All the means $(\bar{d}_1, \bar{d}_2 \ldots \bar{d}_k)$ are concatenated to form a new feature vector called the 'K-spectrum'. Fig. 7.8 shows how the K-inhom function works in mammogram images with the proposed procedures and generates the K-spectrum.

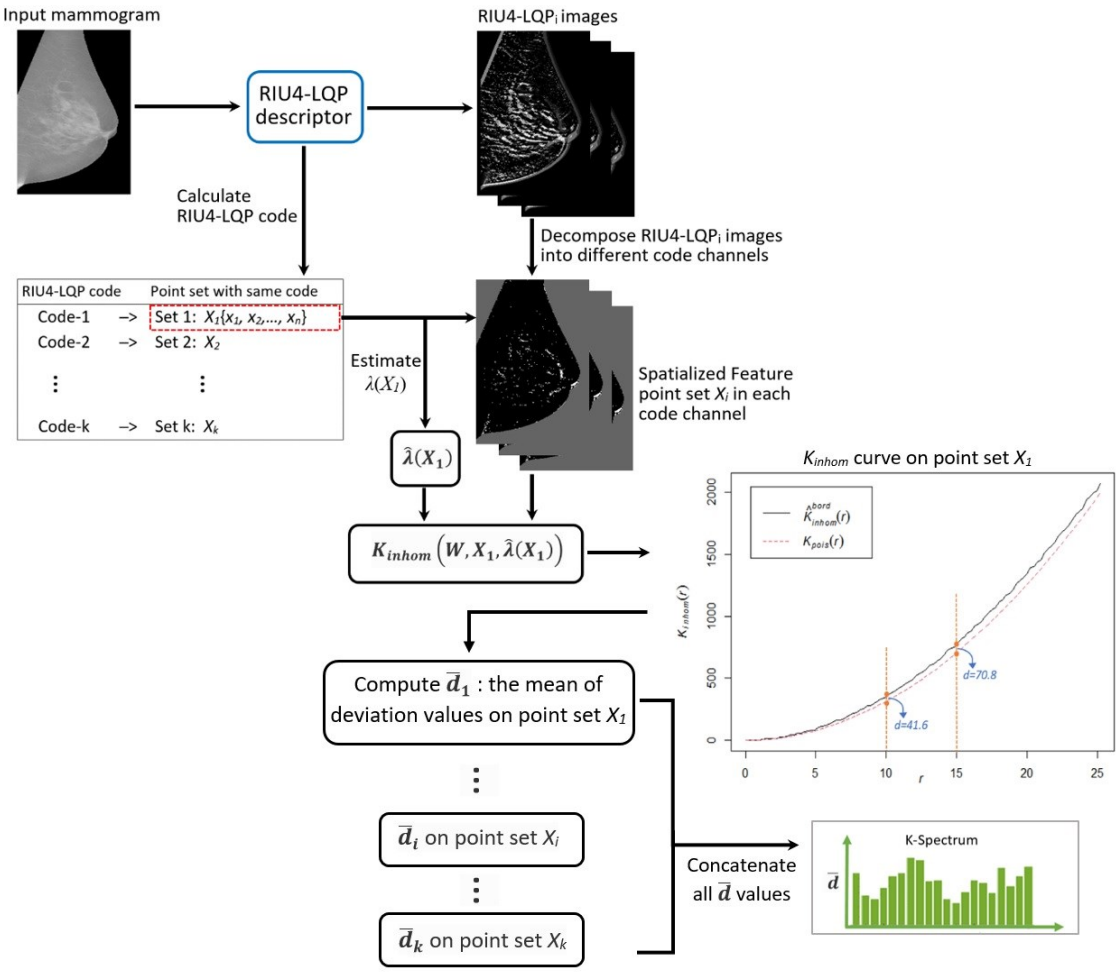


Figure 7.8: Illustration of using K-inhom method and RIU4-LQP operator to generate the new feature vector (K-spectrum).

7.1.4 Feature Concatenation

As discussed in Sections 7.1.1 and 7.1.3, the RIU4-LQP based histograms and the proposed K-spectra are obtained by extracting quantitative and spatial information of feature points. As shown in Fig. 7.1, the classification model concatenates the histograms and K-spectra to form an initial feature space which can be used in the following procedures. Note that histogram information is extracted based on RIU4-LQP images and K-spectrum information is extracted from the spatial distributions of feature points in the decomposed RIU4-LQP code channels. Therefore, they are not in a uniform value range, and a normalization step is used before the features are concatenated. In this work, we use the min-max normalization method to process histograms and K-spectra separately, transforming feature values into the range of [0, 1] for each of the two parts. Concretely, the minimum value of that feature gets transformed into a 0 and the maximum value gets transformed into a 1; every other value gets transformed into a decimal between 0 and 1 as follows.

$$Normalized\ feature value = \frac{value - \min}{\max - \min} \tag{7.7}$$

After doing the normalization process for the histograms and K-spectra, both of them are in the same value range, and then the two parts are concatenated to construct a new feature vector.

7.1.5 Feature Selection

The feature concatenation step concatenates histograms and K-spectra together to produce the initial feature vector. Every image feature in the feature space can play different roles in the classification task; therefore, a feature selection step is necessary to optimise the initial feature set. This study investigates three feature selection methods and gives a comparative analysis under the same experimental conditions.

Dominant Patterns Set (DPS)

The DPS method is proposed in (Guo, Zhao, & Pietikäinen, 2012) to construct a subset of the initial feature set for filtering the most frequently occurring feature patterns. This feature selection method has been used in (Mehta & Egiazarian, 2016)(Rampun, Scotney, Morrow, Wang, & Winder, 2018) (Rampun, Morrow, Scotney, & Wang, 2020) and our proposed classification model in Chapter 6 for compressing feature dimensionality. In DPS, a set of dominant patterns of an image is

defined as a minimum set that can cover $n\%$ ($0 < n < 100$) of all patterns. The formula of filtering dominant patterns has been given in Eq. (6.11) in Chapter 6.

Recursive Feature Elimination (RFE)

RFE method (Guyon, Weston, Barnhill, & Vapnik, 2002) recursively removes features and builds a model on the remaining features. The model accuracy is used to identify which features contribute more than others for predicting target classes. The estimated best feature is assigned a rank score '1', and the least related features have the lowest rank sores.

Feature Importance Ranking (FIR)

The FIR method (Breiman, 2001) uses ensembles of decision trees (e.g. random forest) to compute the relative importance of each attribute. An importance score is given for each feature to indicate that related features play more important roles in the class prediction.

All feature selection methods produce a new sequence of features ranked according to their relevance/importance. The top (best) N features in this sequence can be selected and used to test the classification performance with the training set, and the N-feature set with the highest accuracy is the final feature vector used for the test set.

7.1.6 Classification

Related work in the literature (Tzikopoulos, Mavroforakis, Georgiou, Dimitropoulos, & Theodoridis, 2011)(Oliver et al., 2015) (Rampun, Scotney, Morrow, Wang, & Winder, 2018) (Rampun, Morrow, Scotney, & Wang, 2020) reported their best classification results by using SVM to predict target labels of mammographic density. Therefore, SVM is used in this study for training the classification model and producing classification results on test images. Since this classification work aims to classify mammograms into multiple density categories (3 or 4), a multiclass SVM which is implemented by a one against all (OAA) method is used in this model. To obtain the optimal classification results, the three mostly commonly used kernels (Du, Liu, & Xi, 2015), RBF, Poly, and Sigmoid, are tested in this work. A grid-searching method is used to find the best combinations of parameters (gamma, C, and degree).

7.2 Parameter Optimisation

The methods used in the proposed classification model involve a series of parameters which affect the classification results differently. This section summarises the relevant parameters in different processing steps and addresses the test and optimisation methods.

7.2.1 Relative Parameters

The parameters and the range of their initial values are listed in Table 7.1. In the pre-processing stage, a scaling parameter s is selected from the set of {1, 0.5, 0.25, 0.125, and 0.0625} by testing different mammogram datasets. Comparative analysis is conducted in previous work (Chapter 6) and the same settings of s are used. When using RIU4-LQP to extract the feature set, a multi-scale strategy is used for capturing richer image representation. Referring to the related work in (Rampun, Scotney, Morrow, Wang, & Winder, 2018), three pairs of (R, P) with corresponding settings in Table 7.1 are used. The LQP based method needs an extra two threshold values $\{\tau_1, \tau_2\}$ in its encoding system, which requires manual determination. An automatic approach is proposed in (Peng et al., 2017) by considering the central pixel's intensity (I_c). Similarly we introduce the following empirical rules for adaptively deciding τ_1 and τ_2: $\tau_1 = I_c \times 2\%$ *and* $\tau_2 = I_c \times 7\%$. For generating the K-spectrum feature vector, the K_{inhom} curve is output based on a distance measure (r). However, the range of valid r values varies for different images, as the area of the region of interest (W) depends on the segmented breast region, which does not have a uniform size between different mammograms. After comparing all the observed r-ranges on the whole dataset of images, a maximum valid r-range ([1, 25]) is found and used to generate the K-spectra. In the feature selection step, as the concrete values of (R_i, P_i) of RIU4-LQP are designated in Table 7.1, the feature dimensionality (n_1) of the initial feature set can be obtained by Eq. (6.9).

Table 7.1 Different stages of the classification algorithm and corresponding parameters

Method	Related Parameters
Image pre-processing	1. Resizing scale s = 0.125 (on INbreast) and 0.5 (on MIAS) 2. Median filter size: 3 × 3
RIU4-LQP	In multi-scale method, three pairs of (R, P) are used: 1. (R_1 = 2, P_1 = 10), (R_2 = 4, P_2 = 14), (R_3 = 8, P_3 = 18) 2. Threshold values $\{\tau_1, \tau_2\}$ are decided by an adaptive method
K-inhom function	1. Valid r-range: [1, 25]
Features selection	Initial features number: 1. n_1 = 656 (Histogram or K-spectrum features) 2. n_2 = 1312 (Concatenated features)
SVM classifier	For grid-searching: 1. Kernel candidates: RBF/Poly/Sigmoid 2. γ searching range: [10^{-4}, 10^{3}] 3. C searching range: [10^{-3}, 10^{4}] 4. Degree (only Poly kernel) searching range: {1, 2, 3, 4, 5, 6}

7.2.2 Selection of r-range in K-spectrum

As shown in Table 7.1, the maximum valid r value range [1, 25] is used when applying Baddeley's K-inhom method in mammograms. Since there is no guarantee that the maximum range is the most effective one for K-spectrum features, we narrow this maximum r-range by 5-unit intervals, and 5 sub-ranges are generated and tested for obtaining the optimal K-spectrum. Fig. 7.9 shows that the highest classification accuracy (CA) is 0.83 and the highest AUCROC is 0.95 obtained by using the r-range of [1, 10] on the INbreast dataset. Therefore, we use this r value range as the optimal distance measurement for producing the K-spectrum in the following experiments.

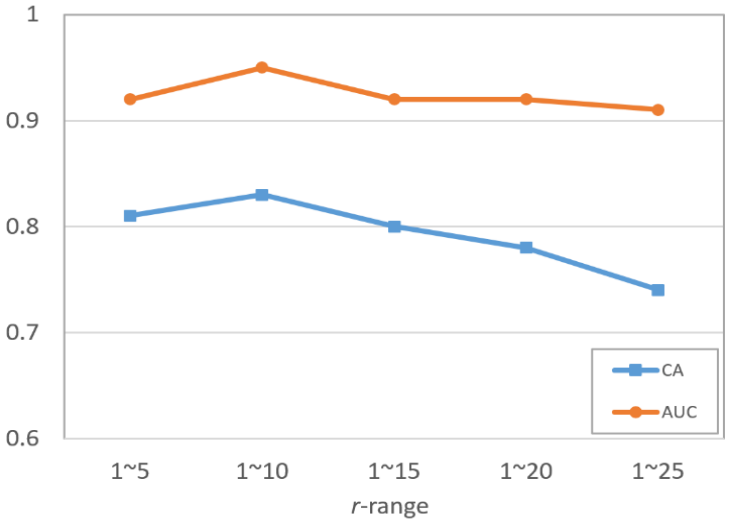


Figure 7.9: Comparison of classification performance on INbreast training set by using different value ranges of distance measurement *r*.

7.2.3 Grid-searching Results for SVM Classifier

We considered three kernels (RBF, Sigmoid, and Poly) and different value ranges of other parameters (gamma, C and degree) for the SVM classifier as given in Table 7.1. The best combination of the kernel and parameters are found by grid-searching on two datasets (Fig. 7.10).

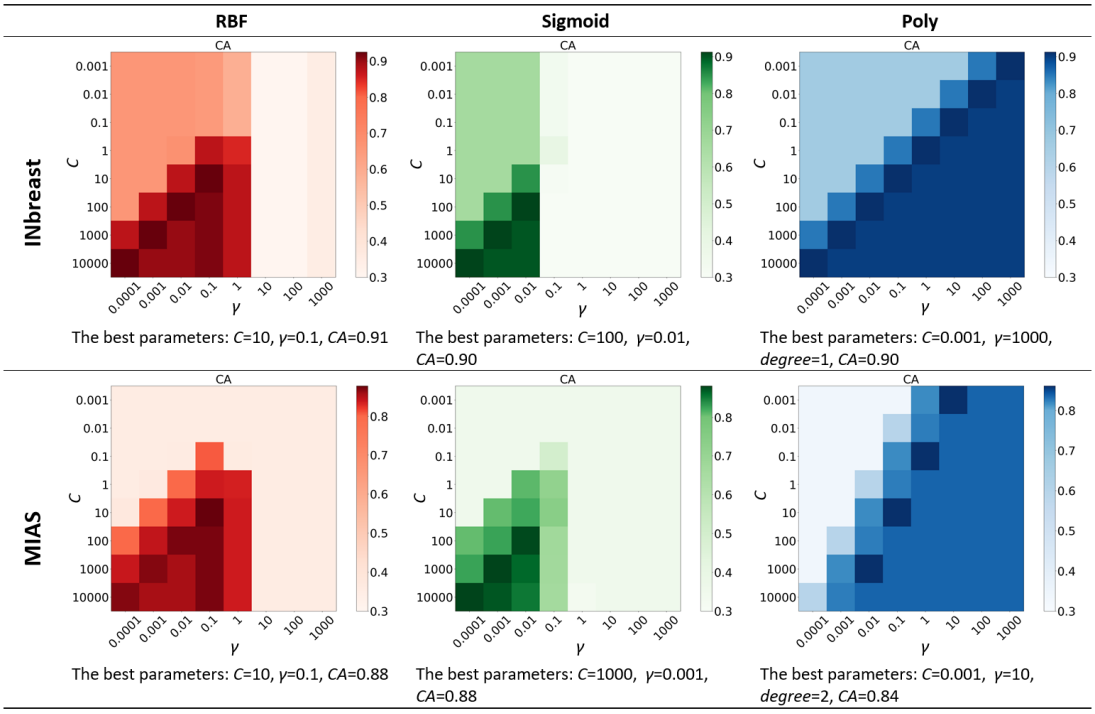


Figure 7.10: Heatmaps of accuracy of grid-searching results using training sets on two datasets.

7.3 Experimental Results and Comparative Analysis

In our experiments, two mammogram datasets, INbreast and MIAS, are used to test the proposed classification model. To give a comprehensive and objective evaluation of the classification performance, different assessment criteria are used. For each test method, classification accuracy (CA) and area under the ROC curve (AUCROC) are calculated as the main performance indices. Since this study investigates the effectiveness of different feature selection methods, the final selected number of features (N) is considered. We also conduct two different test methods: leave-one-woman-out test (Tzikopoulos, Mavroforakis, Georgiou, Dimitropoulos, & Theodoridis, 2011) (Chen, Denton, & Zwiggelaar, 2011) and 10-fold cross validation (Rampun, Scotney, Morrow, Wang, & Winder, 2018) (George & Zwiggelaar, 2019).

7.3.1 Classification Results Using Histogram Information

Histogram information based on RIU4-LQP features is tested first on two datasets. Here, the histogram based feature vector consists of two parts: (i) four split patterns from the RIU4-LQP encoding system with a single local region scale, and (ii) multi-scale patterns by using three pairs of (R_i, P_i) matching to different local regions (Table 7.1). This results in the initial feature dimensionality of 656. Three feature selection methods (section 7.1.5) re-sort the feature vector based on their importance. In each test iteration, a feature subset containing the first N features $\{f_1, f_2 \ldots f_N\}$ ($N \leq 656$) is sent to the classifier for producing the classification result. Then N is increased and the test procedure enters the next iteration, finally obtaining the curves of CA-vs-N. Fig. 7.11 shows the classification results on the two datasets, in which the highest CA is 88.16% and 81.06% on INbreast and MIAS, respectively. Meanwhile, we can also notice that different feature selection methods affect the classification results differently. For INbreast, RFE reduces the feature number N to 161 and obtains the highest AUCROC value (0.95 ± 0.02); for MIAS, the highest AUCROC value is 0.91 ± 0.03 when using FIR with $N = 214$.

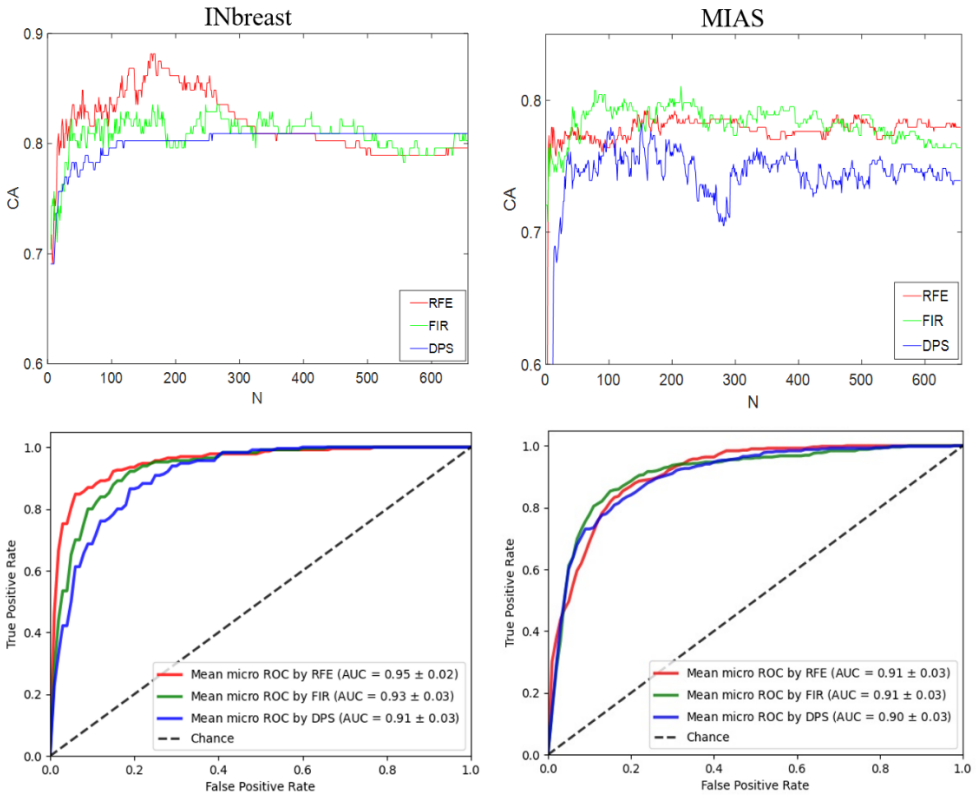


Figure 7.11: Classification accuracy (first row) and AUCROC values (second row) on two datasets using the feature vector based on histograms.

7.3.2 Classification Results Using K-spectrum

K-spectra are extracted in mammograms in the two datasets and used as texture features in the classification model. As the extraction of K-spectra are based on the same RIU4-LQP operator and multi-scale method used for collecting histogram information, the K-spectrum based feature vector also contains 656 features. Classification performance is shown in Fig. 7.12, with the highest CA of 82.89% and 73.60% obtained on the two datasets, respectively.

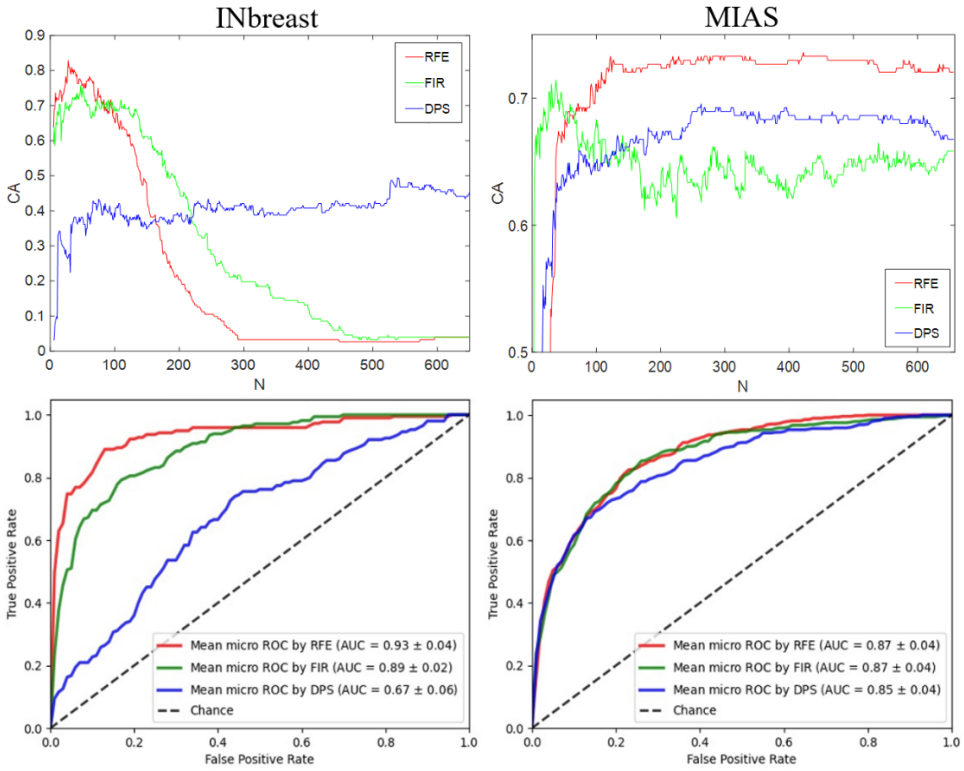


Figure 7.12: Classification accuracy (first row) and AUCROC values (second row) on two datasets using the feature vector based on K-spectra.

7.3.3 Classification Results Using Concatenated Features

The histogram and K-spectrum features are further concatenated to make up a new hybrid feature space. In this step, feature selection procedures become important, as the concatenation operation doubles the feature dimensionality from 656 to 1312 with the relevant parameter settings in Table 7.1. Fig. 7.13 shows the classification results using the hybrid feature vector. Higher CA value of 92.76% on INbreast and 86.96% on MIAS are obtained using RFE when $N = 80$ and $N = 127$, respectively, exceeding the best CA results given in the previous section, using only histogram or K-spectrum features. The AUCROC values are 0.95 ± 0.03 and 0.95 ± 0.02 on the two datasets (Fig. 7.14). Since the classification accuracy is improved significantly on both datasets after combining two feature sets, we can conclude that features extracted from K-spectra can offer complementary image texture features to further improve the classification accuracy.

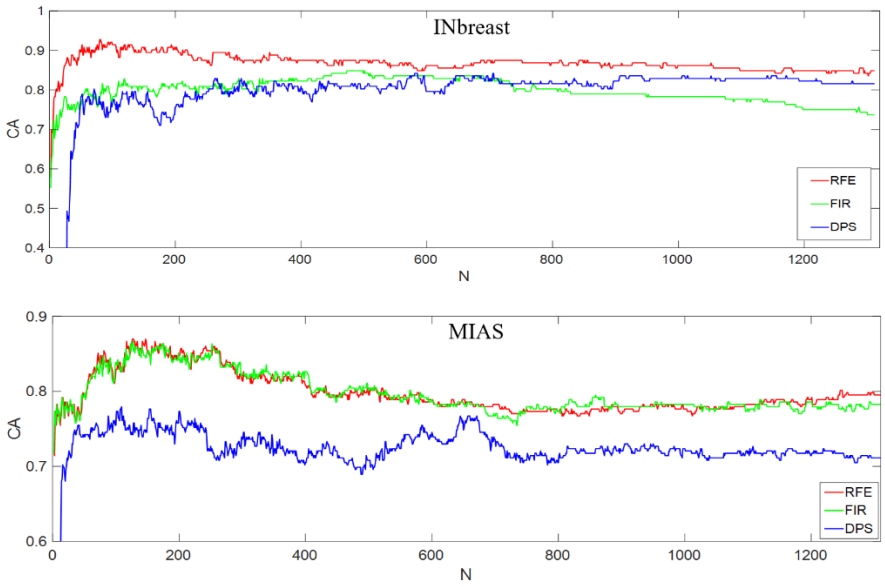


Figure 7.13: Classification accuracy on two datasets using combined texture features.

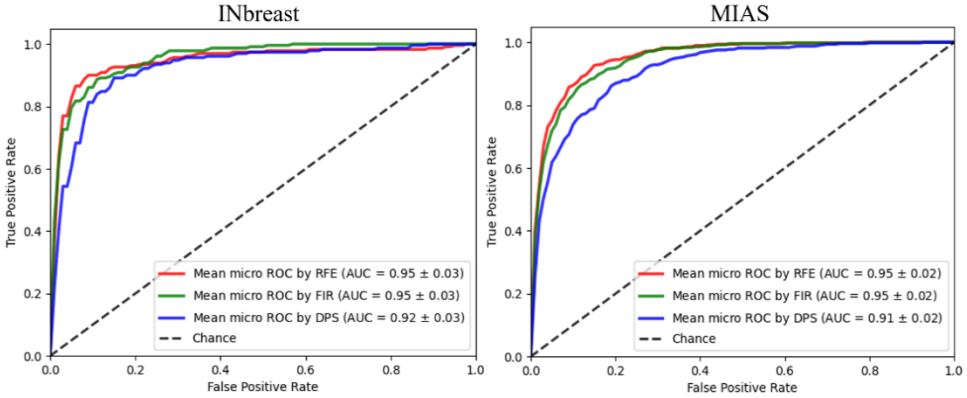


Figure 7.14: AUCROC values on two datasets using different feature selectin methods.

7.3.4 Effect of Feature Selection

Three feature selection methods are used and compared in all experimental analyses in this chapter. In related work (Mehta & Egiazarian, 2016)(Rampun, Scotney, Morrow, Wang, & Winder, 2018), the DPS method was used to optimise texture features for analysing mammograms and other texture images. However, there was no comparative analysis including other feature selection methods. To bridge this gap, we use the selection results by DPS and corresponding classification accuracy as the base line in this work, and the use two other feature selection methods, RFE and FIR, to repeat the feature selection and breast density classification procedures. The comparisons are

given in Table 7.2, from which we can see that RFE works better than the other two methods, with a low number of features used and higher classification accuracy.

Table 7.2 The number (N) of features selected by different methods and corresponding CA values.

Dataset	Feature selection methods	Histograms based feature vector		K-spectra based feature vector		Combined texture features	
		N	Best CA	N	Best CA	N	Best CA
INbreast	RFE	**161**	**88.16**	**28**	**82.89**	**80**	**92.76**
	FIR	164	83.55	47	75.66	473	84.87
	DPS	260	80.92	538	49.34	581	84.21
MIAS	RFE	156	79.19	422	**73.60**	127	**86.96**
	FIR	214	**81.06**	**37**	71.43	127	86.65
	DPS	**152**	78.26	263	69.57	**109**	77.95

7.3.5 Methods Comparison on INbreast Dataset

Different feature extraction methods are compared with each other on the INbreast dataset. As INbreast has not been used widely with the breast density classification task in the literature, by now the available experimental results based on the same dataset are 86% and 80.5% reported in (Rampun, Scotney, Morrow, Wang, & Winder, 2018) (Rampun, Morrow, Scotney, & Wang, 2020), with half images (MLO views) of the dataset tested. We compare 5 progressive transformations of LBP/LQP methods and implement corresponding algorithms to classify INbreast mammograms, performing comparative analysis based on their classification results. We use RIU4-LQP-K and RIU4-LQP-HK to denote texture feature sets based on the K-spectra and the combination of histograms and K-spectra, respectively. Classification performance is evaluated by classification accuracy (CA), AUCROC, Kappa coefficient and F1 score. In addition, a statistical hypothesis test is conducted. The '10-fold cv t-test' method in (Bouckaert, 2003) with a significance level of 0.05 (i.e. alpha = 0.05) is employed between RIU4-LQP-HK and every other method to calculate a *p*-value, which shows the statistical difference between them. Table 7.3 shows that the proposed RIU4-LQP-HK method outperforms other approaches, with the highest CA (91.87 ± 6.28), Kappa (0.8911), and F1 score (0.9216). In the t-test, all other methods present low *p* values (< 0.05), which means that the difference in classification performance is statistically significant.

Table 7.3 Classification performance comparison on INbreast using different Methods.

	CA (%)	AUC (%)	Kappa	F1	*p*-value
RIU2-LBP	78.75 ± 9.35	94.93 ± 1.97	0.7152	0.7953	0.0002
RIU2-LQP	83.75 ± 7.50	94.09 ± 2.19	0.7823	0.8415	0.0002
RIU4-LQP	85.62 ± 12.20	**95.79 ± 2.07**	0.8077	0.8629	0.0251
RIU4-LQP-K	84.38 ± 8.03	95.37 ± 2.02	0.7914	0.852	0.0028
RIU4-LQP-HK	**91.87 ± 6.28**	95.36 ± 2.52	**0.8911**	**0.9216**	–

Furthermore, each subject in INbreast is associated with two different mammogram views (MLO and CC). We therefore divide mammograms into an MLO or CC group along with subject ID information. Classification results are analysed by paired t-test to show the statistical difference between the MLO/CC view groups. If the proposed classification method can work equally well on both mammogram views from the same subject, then the predicted density labels on the two groups are supposed to be the same.

Depending on the proposed classification procedure, each test mammogram is predicted with one density label in {0, 1, 2, and 3} corresponding to the BI-RADS I-IV criteria. The classification performance is evaluated by the predicted category value (mean ± stand deviation), classification accuracy, and the *p*-value in t-test (Table 7.4). The *p*-value is calculated based on the predicted density labels between two groups using a paired t-test. A high *p*-value indicates that the predicted results on the two groups are statically the same, which demonstrates the robustness of the methods. In Table 7.4, we can see that RIU4-LQP and RIU4-LQP-HK present the highest *p*-values, with 1.0 and 0.8855, respectively, and RIU4-LQP-HK produces higher overall CA on both groups than other methods. Therefore, based on the results comparison in Table 7.3 and Table 7.4, we can conclude that the proposed method not only obtains the best classification accuracy, but also is a robust approach by working equally well on MLO/CC view mammograms.

Table 7.4 Classification results comparison and statistical test between INbreast MLO/CC view groups using different methods.

	Predicted category		CA (%)		*p*-value
	MLO group	CC group	MLO group	CC group	
RIU2-LBP	1.37 ± 1.06	1.49 ± 1.14	82.89	75.00	0.5116
RIU2-LQP	1.36 ± 1.07	1.45 ± 1.13	85.53	85.53	0.6091
RIU4-LQP	1.38 ± 1.06	1.38 ± 1.09	85.53	**90.79**	**1.0000**
RIU4-LQP-K	1.42 ± 1.12	1.55 ± 1.08	90.79	75.00	0.4643
RIU4-LQP-HK	1.43 ± 1.10	1.46 ± 1.13	**96.05**	**89.47**	**0.8855**

7.3.6 Methods Comparison on MIAS Dataset

Since MIAS has been a publicly available dataset for some years, a number of research works have been done based on it, including work on breast density classification. We notice that the obtained classification accuracy can be affected by a few factors, such as the number of images used for testing, the results evaluation method, and the target categories. Therefore, when comparing our proposed method with other approaches, these factors are also considered for presenting an objective and fair comparison. In Table 7.5, the CA values are compared between different methods, and we can see that the proposed RIU4-LQP-HK method is very competitive with state-of-the-art approaches, obtaining the highest CA when testing all images in the MIAS dataset.

Table 7.5 Classification results comparison on MIAS dataset.

Methods	Image features and classifier	Image used	Results evaluation method	Results CA (%)	Density categories
(Blot & Zwiggelaar, 2001)	Texture GLCM + KNN	265	Leave-one-image-out	65.0	3 (F, G, D)
(Muhimmah & Zwiggelaar, 2006)	Multiresolution histogram Features + DAG-SVM	321	Leave-one-image-out	77.6	3 (F, G, D)
(Subashini, Ramalingam, & Palanivel, 2010)	Statistical features + SVM	43	3-fold cross validation	95.6	3 (F, G, D)
(Tzikopoulos, Mavroforakis, Georgiou, Dimitropoulos, & Theodoridis, 2011)	Intensity based features + SVM	322	Leave-one-image-out	85.7	3 (F, G, D)
	Intensity based features + KNN		Leave-one-image-out	78.6	
	Intensity based features + SVM		Leave-one-woman-out	77.0	
	Intensity based features + KNN		Leave-one-woman-out	76.4	
(Mario, Mislav, & Krešimir, 2012)	GLCM + KNN	322	Leave-one-image-out	82.0	3 (F, G, D)
(Rampun, Scotney, Morrow, Wang, & Winder, 2018)	LQP + SVM	322	10-fold cross validation	86.13	4 (BI-RADS)
(George & Zwiggelaar, 2019)	LBP + Bayesian Network	321	10-fold cross validation	69.4 ± 0.92	3 (F, G, D)
	ELBP + Bayesian Network			75.4 ± 1.05	
	u-ELBP + Bayesian Network			73.3 ± 0.64	
	M-ELBP + Bayesian Network			77.4 ± 1.06	
	LDP + Bayesian Network			76.0 ± 0.96	
(Rampun, Morrow, Scotney, & Wang, 2020)	RIU2-LBP + SVM	322	5-fold cross validation	73.8 ± 10.6	4 (BI-RADS)
	RIU2-LTP + SVM			81.0 ± 9.5	
	RIU2-LQP + SVM			82.1 ± 7.1	
	RIU2-LSP + SVM			83.3 ± 8.8	
Ours	RIU4-LQP-HK + SVM	322	Leave-one-woman-out	**86.96**	3 (F, G, D)
	RIU4-LQP-HK + SVM	322	10-fold cross validation	**90.61 ± 4.78**	3 (F, G, D)

7.3.7 Robustness of The Proposed Model

As introduced in Section 7.1, the proposed model explores the LQP-based texture descriptor and the K-inhom function to classify the breast density. The parameters used in these methods and their values tested in experiments are listed in Table 7.1. The experimental results displayed in Chapter 6 and Chapter 7 show that different parameters used in the proposed model affect the classification accuracy differently.

Firstly, we focus on the use of the texture descriptor. Two main variants of the conventional LQP method, RIU2-LQP and RIU4-LQP, are developed in Chapter 6. The difference between the two variants is the values used for the parameter T which is

defined as the number of spatial transitions in binary patterns (Section 6.2.2). Specifically, for RIU2-LQP, we set $T \in \{0, 2\}$, and for RIU4-LQP, $T \in \{0, 2, \text{and } 4\}$. Experimental results in Table 6.2 and Table 6.3 show that RIU4-LQP produces approximately 5% higher accuracy than using the RIU2-LQP. For example, an accuracy of 82.50% (± 8.75%) is obtained on INbreast dataset using the RIU4-LQP, while it decreases to 77.50% (± 9.35%) when using the RIU2-LQP, which indicates the superiority of the improved RIU4-LQP method. Meanwhile, the proposed model is robust when working across two different datasets (INbreast and MIAS). Similar accuracy results are obtained (82.50 ± 8.75% vs 80.30 ± 4.55%) when using the RIU4-LQP on the two datasets respectively, and the classification accuracies are 77.50 ± 9.35% vs 74.24 ± 8.60% using the RIU2-LQP.

Secondly, this study explores the spatial distribution information to further improve the classification accuracy in Chapter 7. Baddeley's K-inhom function (Section 7.1.2) is used to extract spatial information based on feature points. The parameter *r* is a distance measure used in the K-inhom function and a value range of *r* (i.e. *r*-range) is used to calculate the K-inhom curve and the K-spectrum feature (Section 7.1.3). A maximum valid *r*-range (i.e. [1, 25]) is found, and it is decomposed into 5 different sub-ranges ([1, 5], [1, 10], [1, 15], [1, 20], and [1, 25]) which are tested in experiments. Experimental results in Fig. 7.9 show that the classification accuracy varies from 74% to 83% by using different *r*-ranges, and the highest accuracy is obtained when *r*-range is [1, 10].

In addition, this study uses three different feature selection methods to compare the corresponding accuracy results. Experimental results in Section 7.3 show that the RFE method produces the highest accuracy for both mammogram datasets, and the parameter *N* which denotes the number of features, is tested in a wide range (i.e. [20, 1200]) to see the difference on the accuracy results. Experimental results in Fig. 7.13 and Table 7.2 show that the accuracy varies from 75% to 93% on INbreast dataset using different *N*, and from 76% to 87% on MIAS dataset. However, the highest accuracy is obtained by using different *N* on two datasets. For INbreast, the accuracy reaches 93% when $N = 80$, and for MIAS, the accuracy of 87% is obtained when $N = 127$.

Through the discussion above, we can conclude that the proposed breast density classification model is robust based on the experimental results on two different datasets, as the highest classification accuracy is obtained on both datasets using this model comparing to other methods. But we also notice that the robustness relies on the

proper choice of relative parameters. The proposed model still is sensitive towards some parameters, for example, the parameter T can cause as much as 5% difference on the accuracy, and the use of the parameter r-range can bring around 10% difference. Some further work will be planned to test the robustness of the proposed model when more mammogram datasets are available for this classification work.

7.4 Chapter Summary

This chapter presents a breast density classification model with novel image texture feature extraction methods. Based on the proposed texture descriptor RIU4-LQP, this study employs Baddeley's K-inhom function to capture spatial distribution information of feature points and construct a new feature vector known as the K-spectrum. After concatenating histogram and K-spectrum information, this chapter also investigates three different feature selection schemes for optimising the initial feature set and improving the classification result. An SVM classifier is trained and used to predict the density labels for test images. Classification results are evaluated using classification accuracy and AUCROC value, and statistical analysis is conducted between different methods. Two mammogram datasets, INbreast and MIAS, are used to test the proposed classification model. Experimental results demonstrate that the proposed method extracts robust and effective texture features in mammograms, which improve the classification performance significantly. Comparing with the state-of-the-art methods, the classification accuracy by using the proposed model is competitive, reaching the best CA of 92.76% and 86.96% on INbreast and MIAS datasets, respectively.

8 CONCLUSIONS AND FUTURE WORK

8.1 Conclusions

This thesis proposes novel algorithms for processing and analysing mammogram images for two research tasks: microcalcification detection and mammographic density classification. The two tasks play important roles in detecting imaged characteristics related to breast cancer diagnosis and risk assessment respectively. For the first work, this thesis develops a complete microcalcification detection framework containing two stages: microcalcification detection and false positive reduction. This thesis places importance on analysis of texture features associated with microcalcificaitons, and proposes novel texture feature enhancement methods and a patch-wise convolutional neural network for accurately detecting positions of microcalcifications in mammograms. For mammographic density classification work, this thesis develops four different classification models which focus on various image feature extraction techniques, feature optimisation methods and feature selection schemes. All developed algorithms have been tested on publicly available mammogram datasets with ground truth information annotated by breast imaging radiologists. Quantitative evaluation and comparative analysis are conducted based on experimental results to demonstrate the effectiveness and robustness of the proposed models. With the two different research tasks, the main contributions and achievements in this thesis are summarised below:

Microcalcification detection

Stage 1: microcalcification detector (Chapter 4)

- Multifractal analysis is applied to analyse mammograms and used to describe texture features of microcalcifications.
- Adaptive alpha value selection rules are developed for highlighting texture patterns that correspond to microcalcifications.
- A linear structure detector is adopted to distinguish microcalcifications from glandular and fibrous tissue.
- Detection performance is tested using INbreast mammograms, obtaining **80.6%** of sensitivity when the average false positive number per image is **90**.

Stage 2: false positive reduction (Chapter 4)

- A Weber's law based approach is used to develop a new multifractal measure for further improving the enhancement of texture features of microcalcifications.
- A patch-wise convolutional neural network classifier is designed to process original mammogram patches and corresponding alpha patches for classifying suspicious microcalcification spots to a true positive or false positive group.
- Classification performance has reached **90%** of sensitivity and **88%** of specificity, indicating that the false positive number is reduced significantly.
- Combining the two stages to produce the final detection results on INbreast dataset, the sensitivity is **77.9%** when the average false positive number is **under 30** per image.

Mammographic density classification

Model 1: using multifractal spectrum (Chapter 5)

- Multifractal spectra are extracted based on sub-images cropped in breast regions and used to capture texture features of fibro-glandular tissue.
- Intensity histograms of breast regions are analysed by standard deviation and skewness for selecting the optimal multifractal spectrum segments for sub-images classification.
- The Chi-square test statistic is used to measure the similarity between the observed and expected multifractal spectrum, based on which sub-images are classified into fatty or dense tissue groups.

- The proposed model produces a rough segmentation contour of dense tissue areas, which is used to compute percentage density and match the BI-RADS density categories. INbreast dataset is used to test the classification model with 20 mammograms used as training set and 78 mammograms as test set, which obtains the classification accuracy of **83.33%**.

Model 2: cascading alpha histograms and LBP (Chapter 5)

- The Local binary pattern (LBP) method and its variants are used to describe texture features related to fibro-glandular tissue in breast regions.
- Alpha histograms in multifractal analysis are extracted using four different local measures and are cascaded with LBP histograms separately to construct a hybrid feature vector.
- An Autoencoder network and the principal components analysis technique are used to optimise the cascaded feature vector.
- The classification model is tested using the whole INbreast dataset (409 images) with 10-fold cross validation, and the classification accuracy is **80.77%** using only **14** features.

Model 3: a novel texture feature descriptor based on local quinary patterns (Chapter 6)

- This model investigates the local quinary pattern (LQP) method for collecting more details of texture features related to breast density classification.
- The basic LQP operator is extended to rotation invariant uniform LQP (RIU4-LQP) by applying a rotation invariant method and considering a wider range of transition numbers. Comparing with conventional rotation invariant uniform patterns (RIU2), this is the first study to extend the bit-wise transition number from 2 to 4 (RIU4), which not only reduces the high feature dimensionality but also keeps a strong capability of recognizing diverse texture patterns.
- Using the extracted texture features, this classification model is tested using two mammogram datasets, INbreast (409 images) and MIAS (322 images), obtaining classification accuracy of **82.50%** and **80.30%,** respectively.

Model 4: introducting spatial distribution analysis of texture descriptors (Chapter 7)

- This model employs Baddeley's K-inhom method to analyse and extract spatial distribution information of feature points based on the proposed RIU4-LQP feature descriptor.

- Using the feature point sets which are decomposed into different RIU4-LQP code channels, the extracted spatial distribution characteristics are used to construct a new feature vector 'K-spectrum' for classifying breast density.
- Histograms and K-spectra are concatenated to form an initial feature set, and machine learning based feature selection methods are adopted to filter the most important and relevant features into the final feature set.
- All mammogram images in two datasets, INbreast (409 images) and MIAS (322 images), are used to test the classification model, and the highest classification accuracy of 92.76% and 86.96% are obtained on the two datasets, respectively.

8.2 Future Work

This thesis proposes a complete framework to detect microcalcification positions in mammograms, and satisfactory detection accuracy is obtained in experiments. Based on this achievement, some further work is planned below:

1) With the detected positions of microcalcification spots, mammogram patches containing potential microcalcification clusters can be extracted automatically and used for analysing cluster features in the following work. In the literature, most research work collected suspicious mammogram patches by radiologists' manual operations.
2) Using the microcalcifications contained patches, segmentation algorithms can be developed to segment microcalcifications for analysing their morphological features and classifying microcalcification clusters into benign or malignant categories.
3) By collecting a sufficiently large number of mammogram patches, deep learning based models can be constructed to automatically learn image features related to microcalcification clusters and to predict corresponding diagnostic categories (e.g. benign, probably benign, and suspicious).

In addition, this thesis addresses four breast density classification models by developing different texture feature extraction methods and feature optimisation approaches. Experimental results and quantitative analysis demonstrate the superiority of the proposed models. Some future directions are listed as follows:

4) In the proposed classification models (model 3 and model 4), texture features based on LBP/LQP methods have been demonstrated as effective and robust for

describing image features of fibro-glandular tissue in mammograms. However, there are some parameters that need to be set and tested manually. Developing an automated method to decide the optimal combination of (R_i, P_i) values in LBP/LQP methods will be planned, for implementing the multi-scale strategy with the most effective texture features extracted.

5) The current work demonstrates the effectiveness of the proposed feature descriptor RIU4-LQP. In future work, we can extend the maximum transition number to $T = 6$ and $T = 8$ and develop new feature descriptors RIU6-LQP and RIU8-LQP.
6) Currently, there is no deep learning based method used for multi-category mammographic density classification work, due to the limitation of available images for training. Novel data augmentation techniques will be developed to increase the size of the datasets to resolve the problems of insufficient images and the imbalance of image distributions in different density categories.

9 REFERENCES

Ahn, C.K., Heo, C., Jin, H., & Kim, J.H. (2017). A novel deep learning-based approach to high accuracy breast density estimation in digital mammography. In Proceedings of the SPIE Medical Imaging 2017: Computer-Aided Diagnosis, Orlando, FL, USA, vol. 10134. https://doi.org/10.1117/12.2254264

Alasadi, A. H. H., & Al-Saedi, A. K. H. (2017, Apr). A Method for Microcalcifications Detection in Breast Mammograms. Journal of Medical Systems, 41(4). https://doi.org/ARTN 6810.1007/s10916-017-0714-7

American Cancer Society. (2017). Cancer facts and figures, Atlanta, GA.

Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. Computational Linguistics, 34(4), 555-596.

Astley, S. M., Harkness, E. F., Sergeant, J. C., Warwick, J., Stavrinos, P., Warren, R., . . . Evans, D. G. (2018). A comparison of five methods of measuring mammographic density: a case-control study. Breast Cancer Research, 20. https://doi.org/ARTN 1010.1186/s13058-018-0932-z

Baddeley, A. J., Møller, J., & Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. Statistica Neerlandica, 54(3), 329-350. https://doi.org/10.1111/1467-9574.00144

Baddeley, A., & Turner, R. (2000). Practical Maximum Pseudolikelihood for Spatial Point Patterns. Australian & New Zealand Journal of Statistics, 42(3), 283-322. https://doi.org/10.1111/1467-842X.00128

Baddeley, Adrian J., Edge Rubak, & R Turner. (2015). Spatial Point Patterns: Methodology and Applications with R. Chapman & Hall/CRC Interdisciplinary Statistics.

Bai, J., Dai, X., Wu, Q., & Xie, L. (2018). Limited-view CT Reconstruction Based on Autoencoder-like Generative Adversarial Networks with Joint Loss. In Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).

Berg, W. A., Campassi, C., Langenberg, P., & Sexton, M. J. (2000). Breast Imaging Reporting and Data System: inter- and intra-observer variability in feature analysis and final assessment. AJR Am J Roentgenol, 174(6), 1769-1777. https://doi.org/10.2214/ajr.174.6.1741769

Blot, L., & Zwiggelaar, R. (2001). Background texture extraction for the classification of mammographic parenchymal patterns. In Proceedings of the 15th Medical Image Understanding and Analysis Meeting.

Bouckaert, R. R. (2003). Choosing between two learning algorithms based on calibrated tests. In Proceedings of the 20th International Conference on International Conference on Machine Learning.

Bouda, M., Caplan, J. S., & Saiers, J. E. (2016). Box-Counting Dimension Revisited: Presenting an Efficient Method of Minimizing Quantization Error and an Assessment of the Self-Similarity of Structural Root Systems. Frontiers in plant science, 7, 149. https://doi.org/10.3389/fpls.2016.00149

Boyd, N. F., Byng, J. W., Jong, R. A., Fishell, E. K., Little, L. E., Miller, A. B., Lockwood, G. A., Tritchler, D. L., & Yaffe, M. J. (1995). Quantitative classification of mammographic densities and breast cancer risk: results from the canadian national breast screening study. Journal of the National Cancer Institute, 87(9), 670-675. https://doi.org/DOI 10.1093/jnci/87.9.670

Boyd, N. F., Osullivan, B., Campbell, J. E., Fishell, E., Simor, I., Cooke, G., & Germanson, T. (1982). Mammographic Signs as Risk-Factors for Breast-Cancer. British Journal of Cancer, 45(2), 185-193. https://doi.org/DOI 10.1038/bjc.1982.32

Brandt, K. R., Scott, C. G., Ma, L., Mahmoudzadeh, A. P., Jensen, M. R., Whaley, D. H., . . . Vachon, C. M. (2016). Comparison of clinical and automated breast density measurements: implications for risk prediction and supplemental screening. Radiology, 279(3), 710-719. doi:10.1148/radiol.2015151261

Braverman, B., & Tambasco, M. (2013). Scale-Specific Multifractal Medical Image Analysis. Computational and Mathematical Methods in Medicine. https://doi.org/Unsp 262931 10.1155/2013/262931

Breast Cancer Foundation New Zealand. https://www.breastcancerfoundation.org.nz/breast-awareness/breast-cancer-facts/breast-cancer-in-nz, last accessed 2019/1/30.

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Byng, J. W., Boyd, N. F., Fishell, E., Jong, R. A., & Yaffe, M. J. (1994). The quantitative-analysis of mammographic densities. Physics in Medicine and Biology, 39(10), 1629-1638. https://doi.org/Doi 10.1088/0031-9155/39/10/008

Cea, M. V. S. d., Nishikawa, R. M., & Yang, Y. (2017). Estimating the Accuracy Level Among Individual Detections in Clustered Microcalcifications. IEEE Transactions on Medical Imaging, 36(5), 1162-1171. doi:10.1109/TMI.2017.2654799

Chen, Z., Denton, E., & Zwiggelaar, R. (2011). Local feature based mammographic tissue pattern modelling and breast density classification. In Proceedings of the 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI).

Ciatto, S., Bernardi, D., Calabrese, M., Durando, M., Gentilini, M. A., Mariscotti, G., . . . Houssami, N. (2012). A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification. Breast, 21(4), 503-506. https://doi.org/10.1016/j.breast.2012.01.005

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37-46. doi:10.1177/001316446002000104

Domingues, I., & Cardoso, J. S. (2014). Using Bayesian surprise to detect calcifications in mammogram images. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 1091-1094.

Du, S., Liu, C., & Xi, L. (2015). A Selective Multiclass Support Vector Machine Ensemble Classifier for Engineering Surface Classification Using High Definition Metrology. Journal of Manufacturing Science and Engineering, 137(1). doi:10.1115/1.4028165

Falconer K. (2005). Random Fractals. Fractal Geometry: Mathematical Foundations and Applications, Second Chichester. UK: John Wiley & Sons.

Fan, R.E., Chen, P.H., Lin, C.J. (2005). Working Set Selection Using the Second Order Information for Training SVM. Journal of Mach. Learning Research 6, 1889–1918.

Fenton, J. J., Abraham, L., Taplin, S. H., Geller, B. M., Carney, P. A., D'Orsi, C., . . . Consortiu, B. C. S. (2011, Aug). Effectiveness of Computer-Aided Detection in Community Mammography Practice. Journal of the National Cancer Institute, 103(15), 1152-1161. https://doi.org/10.1093/jnci/djr206

Gangeh, M. J., Sørensen, L., Shaker, S. B., Kamel, M. S., de Bruijne, M., & Loog, M. (2010). A texton-based approach for the classification of lung parenchyma in CT images. Medical Image Computing and Computer Assisted Intervention (MICCAI).

George, M., & Zwiggelaar, R. (2019). Comparative study on local binary patterns for mammographic density and risk scoring dagger. Journal of Imaging, 5(2). https://doi.org/10.3390/jimaging5020024

George, M., Denton, E., & Zwiggelaar, R. (2018). Mammogram breast density classification using mean-elliptical local binary patterns. Proc. In Proceedings of the 14th SPIE International Workshop on Breast Imaging (IWBI 2018), 107180B. https://doi.org/10.1117/12.2318057

George, M., Rampun, A., Denton, E., & Zwiggelaar, R. (2016). Mammographic ellipse modelling towards birads density classification. International Workshop on Breast Imaging IWDM 2016: Breast Imaging, 423-430. https://doi.org/10.1007/978-3-319-41546-8_53.

Guo, Y., Zhao, G., & Pietikäinen, M. (2012). Discriminative features for texture description. Pattern Recognition, 45(10), 3834-3843. https://doi.org/10.1016/j.patcog.2012.04.003

Guo, Z., Zhang, L., Zhang, D., & Zhang, S. (2010). Rotation invariant texture classification using adaptive LBP with directional statistical features. In Proceedings of the 2010 IEEE International Conference on Image Processing. https://doi.org/10.1109/ICIP.2010.5652209

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. Machine Learning, 46(1), 389-422. https://doi.org/10.1023/A:1012487302797

Hamidinekoo, A., Denton, E., Rampun, A., Honnor, K., & Zwiggelaar, R. (2018). Deep learning in mammography and breast histology, an overview and future trends. Medical Image Analysis, 47, 45-67. https://doi.org/10.1016/j.media.2018.03.006

Han, X. H., Yen-Wei, C. & Gang, X. (2016). Integration of spatial and orientation contexts in local ternary patterns for HEp-2 cell classification. Pattern Recognition Letters, 82, 23-27. https://doi.org/10.1016/j.patrec.2016.02.004

Highnam, R., Brady, M., Yaffe, M. J., Karssemeijer, N., & Harvey, J. (2010). Robust breast composition measurement - Volpara (TM). Digital Mammography, 6136, 342-349.

Hyvarinen, A., Karhunen, J. & Oja, E. (2001). Independent Component Analysis. New York: J Wiley.

Ibrahim, M., & Mukundan, R. (2014). Multifractal Techniques for Emphysema Classification in Lung Tissue Images. 3rd International Conference on Future Bioengineering (ICFB), 115-119.

Ibrahim, M., & Mukundan, R. (2015). Cascaded Techniques for Improving Emphysema Classification in CT Images. Artificial Intelligence Research, 4(2), 112-118. https://doi.org/Doi:10.5430/air.v4n2p112

J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, et al. (2015). Mammographic image analysis society (MIAS) database. https://www.repository.cam.ac.uk/handle/1810/250394, (Accessed 1 June 2020).

Jean-Michel, F., Eric, M., & Florence, P. (2018). Spatial distribution of points. Handbook of Spatial Analysis, 71-111. Insee-Eurostat.

Kallenberg, M., Petersen, K., Nielsen, M., Ng, A. Y., Diao, P. F., Igel, C., . . . Lillholm, M. (2016). Unsupervised Deep Learning Applied to Breast Density Segmentation and Mammographic Risk Scoring. IEEE Transactions on Medical Imaging, 35(5), 1322-1331. doi:10.1109/Tmi.2016.2532122

Kerlikowske, K., Carney, P. A., Geller, B., Mandelson, M. T., Taplin, S. H., Malvin, K., . . . Ballard-Barbash, R. (2000, Dec 5). Performance of screening mammography among women with and without a first-degree relative with breast cancer. Annals of Internal Medicine, 133(11), 855-863. https://doi.org/10.7326/0003-4819-133-11-200012050-00009

Kramer, Mark A. (1991). Nonlinear principal component analysis using autoassociative neural networks. AIChE Journal. 37 (2): 233–243. doi:10.1002/aic.690370209.

Landgrebe, T. C. W., & Duin, R. P. W. (2007). Approximating the multiclass ROC by pairwise analysis. Pattern Recognition Letters, 28(13), 1747-1758. doi:https://doi.org/10.1016/j.patrec.2007.05.001

Lee, J., & Nishikawa, R. M. (2018). Automated mammographic breast density estimation using a fully convolutional network. Medical Physics, 45(3), 1178-1190. https://doi.org/10.1002/mp.12763

Lehman, C. D., Wellman, R. D., Buist, D. S. M., Kerlikowske, K., Tosteson, A. N. A., Miglioretti, D. L., & Surveillance, B. C. (2015, Nov). Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. Jama Internal Medicine, 175(11), 1828-1837. https://doi.org/10.1001/jamainternmed.2015.5231

Li, S. F., Wei, J., Chan, H. P., Helvie, M. A., Roubidoux, M. A., Lu, Y., . . . Samala, R. K. (2018). Computer-aided assessment of breast density: comparison of supervised deep learning and feature-based statistical learning. Physics in Medicine and Biology, 63(2). https://doi.org/ARTN 025005 10.1088/1361-6560/aa9f87

Mario Muštra, Prof. Mislav Grgić & Krešimir Delač. (2012). Breast density classification using multiple feature selection. Automatika, 53(4), 362-372. https://doi.org/10.7305/automatika.53-4.281

McLean, K. E., & Stone, J. (2018). Role of breast density measurement in screening for breast cancer. Climacteric, 21(3), 214-220. https://doi.org/10.1080/13697137.2018.1424816

Mehta, R., & Egiazarian, K. (2016). Dominant rotated local binary patterns (DRLBP) for texture classification. Pattern Recognition Letters, 71, 16-22. https://doi.org/10.1016/j.patrec.2015.11.019

Mohamed, A. A., Berg, W. A., Peng, H., Luo, Y. H., Jankowitz, R. C., & Wu, S. D. (2018). A deep learning method for classifying mammographic breast density categories. Medical Physics, 45(1), 314-321. https://doi.org/10.1002/mp.12683

Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., & Cardoso, J. S. (2012). INbreast: Toward a Full-field Digital Mammographic Database. Academic Radiology, 19(2), 236–248. https://doi.org/10.1016/j.acra.2011.09.01414

Muhimmah, I., & Zwiggelaar, R. (2006). Mammographic density classification using multiresolution histogram information. In Proceedings of the International Special Topic Conference on Information Technology in Biomedicine, Epirus, Greece, 26–28.

Mustra, M., Grgic, M., & Rangayyan, R. M. (2016). Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms. Medical & Biological Engineering & Computing, 54(7), 1003-1024. https://doi.org/10.1007/s11517-015-1411-7

Nanni, L., Lumini, A., & Brahnam, S. (2010). Local binary patterns variants as texture descriptors for medical image analysis. Artificial Intelligence in Medicine, 49(2), 117-125. https://doi.org/10.1016/j.artmed.2010.02.006

Naseem, M., Murray, J., Hilton, J. F., Karamchandani, J., Muradali, D., Faragalla, H., . . . Brezden-Masley, C. (2015). Mammographic microcalcifications and breast cancer tumorigenesis: a radiologic-pathologic analysis. BMC Cancer, 15(1), 307. https://doi.org/10.1186/s12885-015-1312-z

Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. Pattern Recognition, 29(1), 51-59. https://doi.org/10.1016/0031-3203(95)00067-4

Oliver, A., Tortajada, M., Llado, X., Freixenet, J., Ganau, S., Tortajada, L., . . . Marti, R. (2015). Breast density analysis using an automatic density segmentation algorithm. Journal of Digital Imaging, 28(5), 604-612. https://doi.org/10.1007/s10278-015-9777-5

Paskas, M. P., Reljin, B. D., & Reljin, I. S. (2015). Multifractal Techniques for Texture Classification and Image Filtering. 2015 23rd Telecommunications Forum Telfor (Telfor), 791-798.

Paskas, M. P., Reljin, B. D., & Reljin, I. S. (2016, Nov 1). Revision of multifractal descriptors for texture classification based on mathematical morphology. Pattern Recognition Letters, 83, 75-84. https://doi.org/10.1016/j.patrec.2016.01.020

Peng, L., Lin, L., Hu, H., Ling, X., Wang, D., Han, X., & Chen, Y. (2017). Joint weber-based rotation invariant uniform local ternary pattern for classification of pulmonary emphysema in CT images. 2017 IEEE International Conference on Image Processing (ICIP).

Qu, Y., Fu, Q., Shang, C., Deng, A., Zwiggelaar, R., George, M., & Shen, Q. (2020). Fuzzy-rough assisted refinement of image processing procedure for mammographic risk assessment. Applied Soft Computing, 91, 106230. https://doi.org/10.1016/j.asoc.2020.106230

Rampun, A., Morrow, P. J., Scotney, B. W., & Wang, H. (2020). Breast density classification in mammograms: An investigation of encoding techniques in binary-based local patterns. Computers in Biology and Medicine, 122, 103842. https://doi.org/10.1016/j.compbiomed.2020.103842

Rampun, A., Morrow, P. J., Scotney, B. W., & Winder, J. (2017a). Fully automated breast boundary and pectoral muscle segmentation in mammograms. Artificial Intelligence in Medicine, 79, 28-41. https://doi.org/10.1016/j.artmed.2017.06.001

Rampun, A., Morrow, P. J., Scotney, B. W., & Winder, J. (2017b). Breast density classification using local ternary patterns in mammograms. In Proceedings of the Image Analysis and Recognition, Cham. https://doi.org/10.1007/978-3-319-59876-5_51

Rampun, A., Scotney, B. W., Morrow, P. J., Wang, H., & Winder, J. (2018). Breast Density Classification Using Local Quinary Patterns with Various Neighbourhood Topologies. Journal of Imaging, 4(1). https://doi.org/10.3390/jimaging4010014

Reljin, I., Reljin, B., Pavlovic, I., & Rakocevic, I. (2000). Multifractal analysis of gray-scale images. Melecon 2000: Information Technology and Electrotechnology for the Mediterranean Countries, Vols 1-3, Proceedings, 490-493.

Ribli, D., Horvath, A., Unger, Z., Pollner, P., & Csabai, I. (2018, Mar 15). Detecting and classifying lesions in mammograms with Deep Learning. Scientific Reports, 8. https://doi.org/ARTN 416510.1038/s41598-018-22437-z

Ripley, B. (1976). The Second-Order Analysis of Stationary Point Processes. Journal of Applied Probability, 13(2), 255-266. https://doi.org/10.2307/3212829

Saad, G., Khadour, A., & Kanafani, Q. (2016, Dec). ANN and Adaboost application for automatic detection of microcalcifications in breast cancer. Egyptian Journal of Radiology and Nuclear Medicine, 47(4), 1803-1814. https://doi.org/10.1016/j.ejrnm.2016.08.020

Sampat M.P., Bovik A.C., & Whitman G.J. (2008). A model-based framework for the detection of speculated masses on mammography. Med Phys, 35, 2110–2123.

Samuelson, F. W., & Petrick, N. (2006). Comparing image detection algorithms using resampling. In Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro. https://doi.org/10.1109/ISBI.2006.1625167

Scimeca, M., Giannini, E., Antonacci, C., Pistolese, C. A., Spagnoli, L. G., & Bonanno, E. (2014). Microcalcifications in breast cancer: an active phenomenon mediated by epithelial cells with mesenchymal characteristics. BMC Cancer, 14(1), 286. doi:10.1186/1471-2407-14-286

Shi, P., Zhong, J., Rampun, A., & Wang, H. (2018, May 1). A hierarchical pipeline for breast boundary segmentation and calcification detection in mammograms. Computers in Biology and Medicine, 96, 178-188. https://doi.org/10.1016/j.compbiomed.2018.03.011

Sickles, EA, D'Orsi CJ, Bassett LW, et al. (2013). ACR BI-RADS® Mammography. In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA, American College of Radiology.

Singh, B., & Kaur, M. (2018, Mar). An approach for classification of malignant and benign microcalcification clusters. Sadhana-Academy Proceedings in Engineering Sciences, 43(3). https://doi.org/UNSP 3910.1007/s12046-018-0805-2

Slavković-Ilić, M., Gavrovska, A., Milivojević, M., Reljin, I., & Reljin. B. (2016). Breast Region Segmentation and Pectoral Muscle Removal in Mammograms. Telfor Journal, 8(1), 50-55.

Subashini, T. S., Ramalingam, V., & Palanivel, S. (2010). Automated assessment of breast tissue density in digital mammograms. Computer Vision and Image Understanding, 114(1), 33-43. https://doi.org/10.1016/j.cviu.2009.09.009

Tan, X., & Triggs, B. (2010). Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Transactions on Image Processing, 19(6), 1635-1650. https://doi.org/10.1109/TIP.2010.2042645

Tzikopoulos, S. D., Mavroforakis, M. E., Georgiou, H. V., Dimitropoulos, N., & Theodoridis, S. (2011). A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry. Computer Methods and Programs in Biomedicine, 102(1), 47-63. https://doi.org/10.1016/j.cmpb.2010.11.016

U.S. Breast Cancer Statistics. http://www.breastcancer.org/symptoms/understand_bc/statistics, 2016 (last accessed 20 May 2020)

Volpara health. https://volparahealth.com/our-products/volparadensity/downloads/ (last accessed December 2020)

Wang, J., & Yang, Y. Y. (2018). A context-sensitive deep learning approach for microcalcification detection in mammograms. Pattern Recognition, 78, 12-22. https://doi.org/10.1016/j.patcog.2018.01.009

Wang, J., Yang, Y. Y., & Nishikawa, R. M. (2013). Reduction of False Positive Detection in Clustered Microcalcifications. 2013 20th IEEE International Conference on Image Processing (ICIP 2013), 1433-1437.

WHO report. (2019). https://www.who.int/cancer/prevention/

Wolfe J. N. (1976). Risk for breast cancer development determined by mammographic parenchymal pattern. Cancer, 37(5), 2486–2492. https://doi.org/10.1002/1097-0142(197605)37:5

Xue, Y. K., & Bogdan, P. (2017). Reliable Multi-Fractal Characterization of Weighted Complex Networks: Algorithms and Implications. Scientific Reports, 7. doi:ARTN 748710.1038/s41598-017-07209-5

Yelampalli, P.K.R., Nayak, J. & Gaidhane, V.H. (2019). A novel binary feature descriptor to discriminate normal and abnormal chest CT images using dissimilarity measures. Pattern Anal Applic 22, 1517–1526. https://doi.org/10.1007/s10044-018-00771-2

Zhang, E. H., Wang, F., Li, Y. C., & Bai, X. N. (2014). Automatic detection of microcalcifications using mathematical morphology and a support vector machine. Bio-Medical Materials and Engineering, 24(1), 53-59. https://doi.org/10.3233/Bme-130783

Zheng, Y., Keller, B. M., Ray, S., Wang, Y., Conant, E. F., Gee, J. C., & Kontos, D. (2015). Parenchymal texture analysis in digital mammography: A fully automated pipeline for breast cancer risk assessment. Medical Physics, 42(7), 4149-4160. https://doi.org/10.1118/1.4921996

Zwiggelaar, R., Astley, S. M., Boggis, C. R. M., & Taylor, C. J. (2004, Sep). Linear structures in mammographic images: Detection and classification. IEEE Transactions on Medical Imaging, 23(9), 1077-1086. https://doi.org/10.1109/Tmi.2004.828675

10 APPENDICES

APPENDIX 1 A DETAILED DESCRIPTION OF THE DESIGN OF RIU4-LQP

This section details the computation of the '*index*' value in Eq. (6.10):

For a given number of neighbourhood pixels P (e.g. $P = 16$), when the number of bit transitions $T = 4$, there will be exactly two distinct contiguous blocks of 1's in the bit pattern. Let X and Y denote the number of 1's in each of those contiguous segments. We regulate $X \leq Y$, and the following conditions must be satisfied:

(1) $X \geq 1,\ Y \geq 1$

(2) $X + Y \leq P - 2$

All possible '*X*-*Y*' patterns can be summarised within an '*X*-*Y*' pattern square in size of $(P - 3) \times (P - 3)$, where the row number corresponds to the value of X and the column number corresponds to the value of Y (Fig. A.1).

The second part of Eq. (6.10) (when $X = 1$) calculates the '*index*' value in the first row situation in Fig. A.1. The '*index*' value is counted from 1 (i.e. when the '*X*-*Y*' pattern is '1-1'). When $X \geq 2$, Eq. (6.10) calculates the '*index*' value by accumulating other preceding rows' patterns. For example, if the target '*X*-*Y*' pattern is '3-3', $\sum_{n=1}^{X-1}(P-3-2(n-1))$ is used to calculate how many patterns there are in the first two rows and then add its offset position '$Y - X + 1$' in the third row.

By this way, we can also compute the total number of '*X*-*Y*' patterns (i.e. '*index*' value range) when P is designated, which equals to one-quarter area of the square. We already know that when $T \in \{0, 2\}$, the LQP_i^{riu4} code is same with LQP_i^{riu2}, ranging from 0 to P, and all the remaining situations (i.e. $T > 4$) share one unified code. Therefore, the total number of LQP_i^{riu4} patterns in Eq. (6.9), $\lceil (P^2 + 11)/4 \rceil$, is obtained by calculating the '*index*' value range and plus $P + 2$.

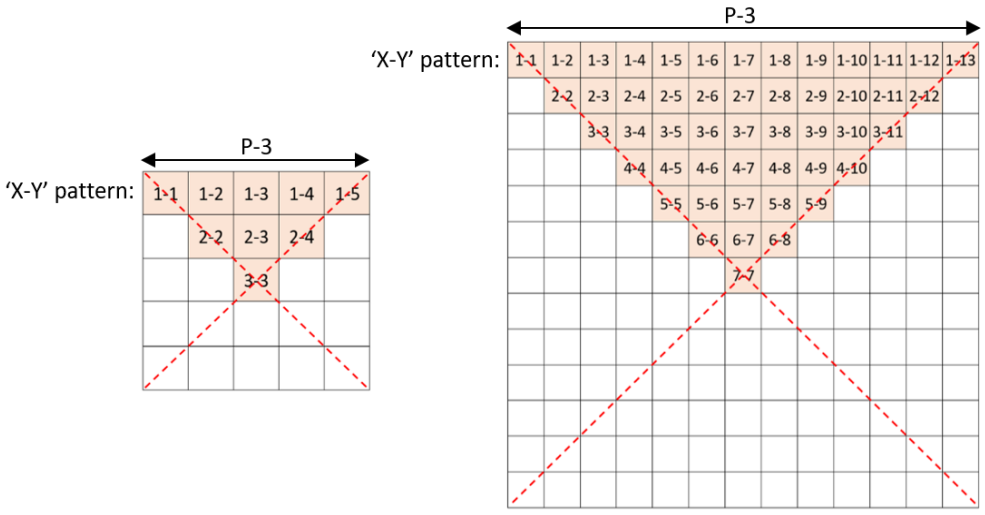


Fig. A.1. Examples of '*X-Y*' pattern map with $P = 8$ and $P = 16$ when $T = 4$.

Fig. A.2 illustrates all possible '*X-Y*' combinations when $P = 8$ and $T = 4$. In the case of $P = 8$, if $T \in \{0, 2\}$, repeating the procedure of LQP_i^{riu2}, 9 pattern codes in range of [0, 8] will match to each pattern; if $T = 4$, then $X + Y \leq 6$, and the parameter '*index*' can be computed and results in [1, 9], which make the final LQP_i^{riu4} code (*index* + *P*) locating in [9, 17]; all the other patterns will be given a uniform code: 18. Similarly, when *P* is increased to 10, the procedure outputs 28 LQP_i^{riu4} patterns; $P = 14$, 52 LQP_i^{riu4} patterns; P=18, 84 LQP_i^{riu4} patterns; and so on.

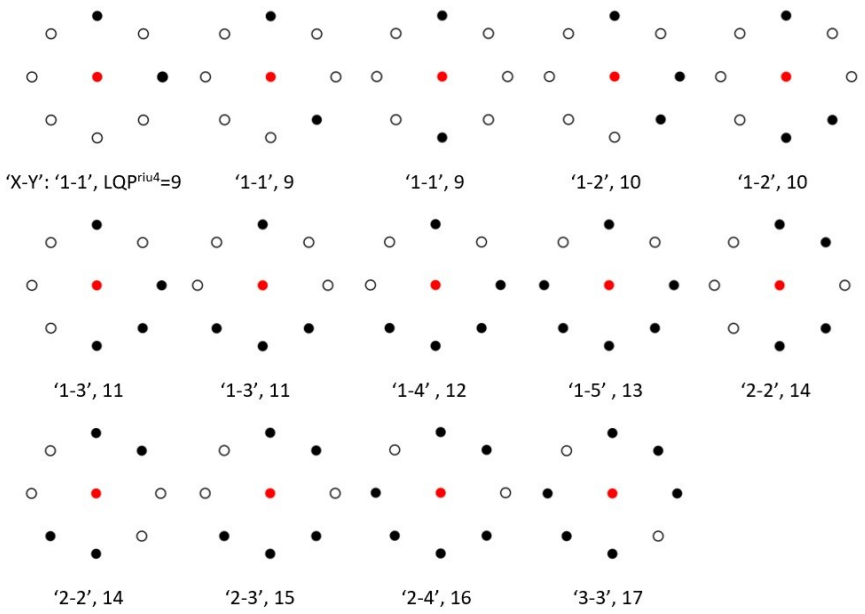


Fig. A.2. All '*X-Y*' patterns and their LQP^{riu4} codes with $P = 8$, $T = 4$.

APPENDIX 2 GLOSSARY

TERM	DEFINITION
Breast Imaging-Reporting and Data System (BI-RADS)	A quality assurance system and lexicon originally designed for reporting mammography results from patients with breast cancer, and published as an atlas by the American College of Radiology. The BI-RADS Atlas is a peer-reporting (i.e., not for lay or patient use) tool which now includes three imaging modalities: mammography, ultrasound and MRI.
Computer Aided Detection	Use of sophisticated computer programs designed to recognize patterns in images.
Craniocaudal (CC) View	A standard view taken from above during routine-screening mammography and during diagnostic mammography. In the CC view, the entire breast parenchyma should be depicted; the fatty tissue closest to the chest wall should appear as a dark strip on the mammogram and behind that, one should see the pectoral muscle.
False Negative Result	A test result that indicates that the abnormality or disease being investigated is not present when in fact it is.
False Positive Result	A test result that indicates that the abnormality or disease being investigated is present when in fact it is not.
Feature Extraction	Feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is related to dimensionality reduction.
Full Field Digital Mammogram (FFDM)	Similar to conventional mammography (screen film mammography) except that a dedicated electronic detector system is used to computerize and display the X-ray information. Compared to SFM, FFDM images generally have a higher resolution and quality.

Histogram	The histogram of an image visualizes the distribution of the brightness in the image by plotting the number of occurrences of each brightness level.
Image Contrast	Any difference in luminance level between regions of interest.
Image Enhancement	Image enhancement is the procedure of improving the quality and information content of original data before processing.
Image Texture	An image texture is a set of metrics calculated in image processing designed to quantify the perceived texture of an image. Image texture gives us information about the spatial arrangement of colour or intensities in an image or selected region of an image.
Malignant Tumour	A tumour that has the potential to become lethal through destructive growth or by having the ability to invade surrounding tissue and metastasize.
Mammogram	X-ray image of the breast.
Mammographic Density	A radiological appearance of the mammary glands during mammography in which the epithelial and stromal elements of the breasts are more prominent than the fatty components. Breast density increases the likelihood that a mammogram will be difficult to interpret and that a patient will subsequently develop breast cancer.
Mammography	Technique for imaging breast tissues with X rays.
Mammography Screening	X-ray-based breast imaging in an asymptomatic population with the goal of detecting breast tumours at an early stage.
Mediolateral Oblique (MLO) View	A standard mammographic view taken from an oblique or angled view, which is the most important projection as it allows imaging of the greatest amount of breast tissue and is preferred over the lateral 90º projection. With the MLO view, the pectoral muscle should be depicted obliquely from above and visible down to the level of the nipple or further beyond that; the muscle should bulge outward as a sign that it is relaxed.

Microcalcifications	Tiny calcium deposits within the breast, singly or in clusters; often found by mammography. They may be a sign of cancer.
Over-diagnosis	Labelling an abnormality as cancer when it in fact is not likely to become a lethal cancer.
Percentage Breast Density	Percentage density, or the fraction (%) of the breast with densities, is the ratio of the dense area to the total breast area.
Scattered Fibro-glandular Tissue	A term used to describe breast tissue that is made up of mostly fatty tissue and also has some dense fibrous tissue and glandular tissue. On a mammogram, the dense areas of the breast make it harder to find tumours or other changes.
Screen Film Mammogram (SFM)	Conventional mammography in which the X rays are recorded on film.